Enhanced Structural and Dynamical Characterization Combining Gold Nanocrystal Labels and Small-Angle X-ray Scattering

Thomas Zettl

München 2018

Enhanced Structural and Dynamical Characterization Combining Gold Nanocrystal Labels and Small-Angle X-ray Scattering

Thomas Zettl

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Thomas Zettl
aus Augsburg

München, den 28.02.2018

Erstgutachter: Prof. Dr. Jan Lipfert
Zweitgutachter: PD Dr. Theobald Lohmüller
Tag der mündlichen Prüfung: 25.04.2018

Zusammenfassung

Die meisten biologische Makromoleküle existieren in der Natur nicht nur in einer bestimmten Konformation sondern nehmen viele verschiedene Formen an. Je nach räumlich angenommener Anordnung können diese Moleküle unterschiedliche Funktionen erfüllen, da durch Anpassung an äußere Einflüsse gezielt aktive Zentren freigelegt, oder unzugänglich gemacht werden können. Eine genauste Untersuchung des Ensembles, der freien Energie und des strukturellen Aufbaus von Molekülen und biologischen Komplexen ist deshalb notwendig, um deren Funktionsweise zu verstehen. Als Beispiele können diverse Motive wie Schleifen, Haarnadelstrukturen oder Fehlpaarungen aus Nukleinsäuren genannt werden, die durch ihr Auftreten, unter anderem in der Genregulation, die Genexpression beeinflussen, indem sie das Binden von Polymerase-Proteinkomplexen entscheidend stören. Außerdem können gezielte Experimente und daraus resultierende Erkenntnisse neue Einblicke in den Prozess der Proteinfaltung liefern, um Ursachen möglicher Fehlfaltungen offenzulegen. Um diese Untersuchungen erfolgreich umzusetzen bedarf es allerdings spezieller Methoden. Im Zuge der vorliegenden Arbeit wurden neue Verfahren entwickelt, um diese Grundlegende Abläufe und deren Ursachen genauer untersuchen zu können. Dafür wurden kleine Goldnanopartikel (~1.4 Nanometer Durchmesser) spezifisch und irreversibel über chemische Bindungen an biologische Makromoleküle gekoppelt und mittels Kleinwinkel-Röntgenstreuung untersucht. Dies geschah in Zusammenarbeit mit den Gruppen von Daniel Herschlag und Pehr A. B. Harbury an der Stanford University.

Zunächst wurde ein Verfahren zur akkuraten Bestimmung von Abständen zwischen zwei Goldnanoteilchen erarbeitet. Diese können paarweise an Molekülen angebracht werden. Die Methode ermöglicht es nicht nur eine feste Distanz, sondern auch eine Verteilung von Abständen, wie bei flexiblen Molekülen üblich, auf einer genormten Skala zu bestimmen. Dafür wurde die anomale Dispersion von Goldnanopartikel verwendet, welche um Absorptionskanten von Atomen auftritt. Dabei steigt der Wirkungsquerschnitt sprunghaft an und Röntgenstrahlen werden vermehrt absorbiert. Aufgrund dieser Auswirkung ist es möglich, den Streubeitrag der zwei Goldpartikel von den restlichen Streutermen zu trennen. Die Wellen, die sich von beiden Goldteilchen beginnend ausbreiten, erzeugen ein charakteristisches Interferenzprofil das mittels Fourier-Transformation in eine genaue Abstandsverteilung umgerechnet werden kann. Deshalb wird diese Methode auch als Röntgenbeugungs-Interferometrie bezeichnet. Um das Verfahren zu benchmarken, wurden Goldnanoteilchen paarweise an gegenüberliegenden Enden von doppelsträngiger Desoxyribonukleinsäure (DNS) (10 Basenpaare-30 Basenpaare) angebracht und vermessen. Die resultierenden Abstände erzielten eine ausgezeichnete Übereinstimmung mit bereits existierenden Untersuchungen und belegen damit die erfolgreiche Anwendung dieser neuen Methode.

Des Weiteren wurde die Holliday-Struktur, ein wichtiges DNS-Motiv das beim Austausch von genetischen Informationen hilft, und deren Konformation in Abhängigkeit der Ionenstärke mittels Röntgenbeugungs-Interferometrie untersucht. Im Folgenden wurden die Abstandsverteilungen zwischen verschiedenen Enden der DNS-Helices unter niedriger, mittlerer und hoher Ionenstärke vermessen. Bei mittlerer Salzkonzentration wurde dabei beobachtet,

dass die Struktur nicht wie bei hohem Salzgehalt zwei symmetrische Konformationen durch $\pi-\pi$–Wechselwirkungen annimmt, sondern sich in einen neuen Zustand organisiert. Dieser Effekt dient in der Natur möglicherweise als Regulierungsprozess, da auf diese Weise die Bindung von Enzymen beeinflusst werden kann.

Erfahrungsgemäß liefert die Auflösung von Kleinwinkel-Röntgenstreuung nur unzureichende Erkenntnisse über die genaue Lage und Ausrichtung von Untereinheiten in einem biologischen Makromolekül, weshalb sich dieser Herausforderung in einem gesonderten Projekt angenommen wurde. Hierzu wurde ein weiteres Verfahren ausgearbeitet, dass die genaue Lokalisierung modifizierter Seitengruppen in niedrig aufgelösten Formrekonstruktionen, welche aus Messdaten durch Kleinwinkel-Röntgenstreuung gewonnen werden, erlauben. Dafür wurden einzelne Goldnanoteilchen an Moleküle mit bekannter Struktur gekoppelt und die experimentell bestimmte Position auf Übereinstimmung mit der vorhergesagten Lage verglichen. Als Modellsysteme wurden verschiedene Desoxyribonukleinsäure-Moleküle, ein Ribonukleinsäure-Konstrukt sowie ein Protein verwendet. Das Verfahren erzielte für alle Moleküle ein hervorragende Übereinstimmung zwischen der rekonstruierten Standorte und der berechneten Positionen.

Preface

Most biological macromolecules typically function through series of conformations. Thus, a complete understanding of macromolecular structures requires knowledge of the ensembles that represent probabilities on a conformational free energy landscape. Moreover, significant conformational changes of molecules can be triggered by external stimuli and are typically integrally involved in the functions of biomolecules. There are many structural motifs such as kinks, hairpins, D-loops, or mismatches in nucleic acids which cause structural folding in special tertiary structures or a broad range of dynamic conformations. As nucleic acids play a key role in genetic recombination, these deformed states can influence the patterning of nucleosomes or affect the regulation of gene expression by altering the binding of transcription factors. Moreover, a precise knowledge of nucleic acid solution ensembles can raise our capacity to design new engineered nanostructures to a new level. Besides, new discoveries in the field of protein conformational ensembles can help to precisely dissect complex folding pathways and lead to a targeted design of new therapeutics.

However, many techniques in structural biology provide information on averaged structures or predominant states of biological macromolecules and their complexes. Thus, considering the importance of structural details and the entire conformational ensemble to understand folding, recognition, and function, tools are needed to probe and quantify the solution structural conformation. While nuclear magnetic resonance (NMR), crystallography and electron microscopy (EM) techniques have provided invaluable details by solving molecular structures or measuring intramolecular distances, they suffer from size limitations or require exceptional environmental conditions. Small-angle X-ray scattering (SAXS), on the contrary, is a powerful technique to probe the structure of biological macromolecules and their complexes under virtually arbitrary solution conditions without the need for crystallization. Moreover, by site-specific labeling of macromolecules with small gold nanocrystals in small-angle scattering experiments, novel information about their structure and dynamics can be gained to provide testable models on the atomic scale.

In this thesis, I show that small gold nanocrystals can be used as markers in SAXS experiments to provide additional information about the conformation of biological macromolecules in solution. I designed a protocol to reliably record X-ray scattering interferometry (XSI) data and developed a graphical user interface (GUI) to allow a rapid analysis of custom diffraction signals. Furthermore, I used XSI to examine the conformational landscape of a DNA Holliday junction motif. Therefore I probed the structural ensemble of a DNA four-way junction under varying solution conditions. Moreover, I combined anomalous small-angle X-ray scattering with X-ray scattering interferometry to record absolute distance distribution of entire molecular ensembles on the atomic scale. I further used single gold nanocrystals to provide a sequence-to-3D structure map in SAXS shape reconstructions.

An introduction to both the theoretical concepts and the experimental implementation of small-angle X-ray scattering is provided in Chapters 1, 2 and 3. The results of this thesis are stated in Chapters 4 to 7. Were each of the chapters provides a separate introduction and the

associated Materials and Methods section.

In Chapter 4, a detailed protocol on how to prepare gold labeled samples and to record the corresponding X-ray scattering interferometry profiles, which was introduced and developed by the groups of Pehr Harbury and Daniel Herschlag. Moreover, a step-to-step guide for a novel custom-written graphical user interface is provided. This chapter is based on Zettl *et al.* [1].

Chapter 5 presents a new method to determine accurate distance distributions using anomalous small-angle X-ray scattering and pairs of gold labels. The method uses the anomalous dispersion of X-rays close to a gold absorption edge to separate the gold-gold interference pattern from other scattering terms. The results are benchmarked against distance distributions obtained from native X-ray scattering interferometry of nucleic acid constructs. This chapter is adapted from Zettl *et al.* [2].

Chapter 6 focuses on the characterization of structural conformation of a DNA Holliday junction motif. The conformational ensemble is probed by Förster resonance energy transfer and X-ray scattering interferometry to study the conformational ensemble of the Holliday junction under a wide range of solution conditions. In addition, novel findings on the structure for individual Holliday junction states are presented. We used these experimental data to generate and test physical models of Holliday junctions to demonstrate the change of the energy landscape under different counterion environments.

Chapter 7 demonstrates that single gold nanocrystal labels attached to specific residues of biological macromolecules can serve as fiducial markers and thus provide a map of the primary sequence (and therefore domain structure) to the low-resolution 3D electron density map computed from SAXS data. We show that our approach is broadly applicable by evaluating its performance on gold-labeled DNA and RNA constructs and extend its application to gold-labeled proteins using a novel labeling approach. This chapter is adapted from Zettl *et al.* [3].

SAXS measurements and method development were accomplished in collaboration with the group of Daniel Herschlag and Pehr A. B. Harbury at Stanford University.

A conclusion and outlook for future research projects are presented in Chapter 8.

Publications

The results presented in this thesis were published in the form of three peer-reviewed publications:

1. Thomas Zettl, Rebecca S. Mathew, Sönke Seifert, Sebastian Doniach, Pehr A. B. Harbury[†] and Jan Lipfert[†], **'Absolute Intramolecular Distance Measurements with Ångström–Resolution Using Anomalous Small–Angle X-ray Scattering.'**, *Nano Letters* 15(5), 5353–5357 (2016).
   DOI:10.1021/acs.nanolett.6b01160
   [†] Joint corresponding authorship.

2. Thomas Zettl, Rhiju Das, Pehr A. B. Harbury[†], Daniel Herschlag[†], Jan Lipfert[†], Rebecca S. Mathew and Xuesong Shi, **'Recording and Analyzing Nucleic Acid Distance Distributions with X–ray Scattering Interferometry (XSI).'**, *Current Protocols*, in Press.
   [†] Joint corresponding authorship.

3. Thomas Zettl, Rebecca S. Mathew, Xuesong Shi, Sebastian Doniach, Daniel Herschlag, Pehr A. B. Harbury and Jan Lipfert, **'Gold Nanocrystal Labels Provide a Sequence–to–3D Structure Map in SAXS Reconstructions .'**, *Science Advances*, 4, eaar4418, (2018).
   DOI:10.1126/sciadv.aar4418

I have further contributed to the following peer-reviewed publication, which has not been included as main result in this thesis:

4. Kerstin Göpfrich[*], Thomas Zettl[*], Anna E. C. Meijering, Silvia Hernández-Ainsa, Samet Kocabey, Tim Liedl, and Ulrich F. Keyser, **'DNA-tile structures induce ionic currents through lipid membranes.'**, *Nano Letters* 15(5), 3134–3138 (2015).
   DOI:10.1021/acs.nanolett.5b00189
   [*] Authors contributed equally to this work.

Contents

Part I

Introduction

Chapter 1
Introduction

Many functional properties of biological macromolecules, such as nucleic acids or proteins, are given by the structural conformation under certain environmental conditions to regulate processes within living organisms or artificial machineries. Thus, obtaining structural insights is key to understand the fundamental interaction between multiple components and the molecular dynamics occurring in all biological systems. Nowadays, structural insights are obtained by a wide range of techniques, such as X-ray crystallography, solution based nucleic magnetic resonance (NMR) or small-angle X-ray scattering (SAXS), cryo-electron microscopy (EM) and atomic force microscopy (AFM) imaging. Typically, very high resolutions can be obtained using X-ray crystallography, NMR or cryo-EM. Those reveal details on the atomic scale [4–9] whereas AFM imaging and SAXS only provide a resolution of ~1 nm for soft matter probes [10–16]. Moreover, the techniques differ in environmental conditions such as the need for high quality crystals in X-ray crystallography, plunge-frozen samples in cryo-EM, or molecules fixed on a substrate as for AFM imaging or a copper grid in case of vacuum electron microscopy.

SAXS was discovered by A. Guinier in the 1930s during his work on metal alloys [17]. From this moment on, SAXS was developed to become a powerful method for the structural characterization from small-angle scattering patterns of non-crystalline samples of a broad range of materials starting from anorganic compounds such as metals or salts and ranging to biological macromolecules. Moreover, the advances in the theoretical understanding and the development of increasingly powerful X-ray sources since the 1970s allowed for exciting applications and fascinating discoveries.

Similar to other techniques for structural characterization, SAXS requires small amounts of highly purified samples desolved in aqueous solution. Furthermore, the improvement in data collection and analysis combined with high throughput setups allow for rapid screening of samples under a wide range of solution conditions. The obtained data can almost immediately be processed and transfered into low-resolution electron density maps depicting the outer particle shape or other molecular parameters (see Fig. 1.1).

Like X-ray crystallography, SAXS is based on coherently scattered X-ray waves. However, the recorded scattering pattern is not based on interfering waves scattered by atoms in a crystalline lattice. Instead, the incoming focused beam of X-rays is scattered off its initial trajectory by particles randomly oriented in solution. It is always required to record the scattering profile of the macromolecule in solution as well as the corresponding buffer solution.

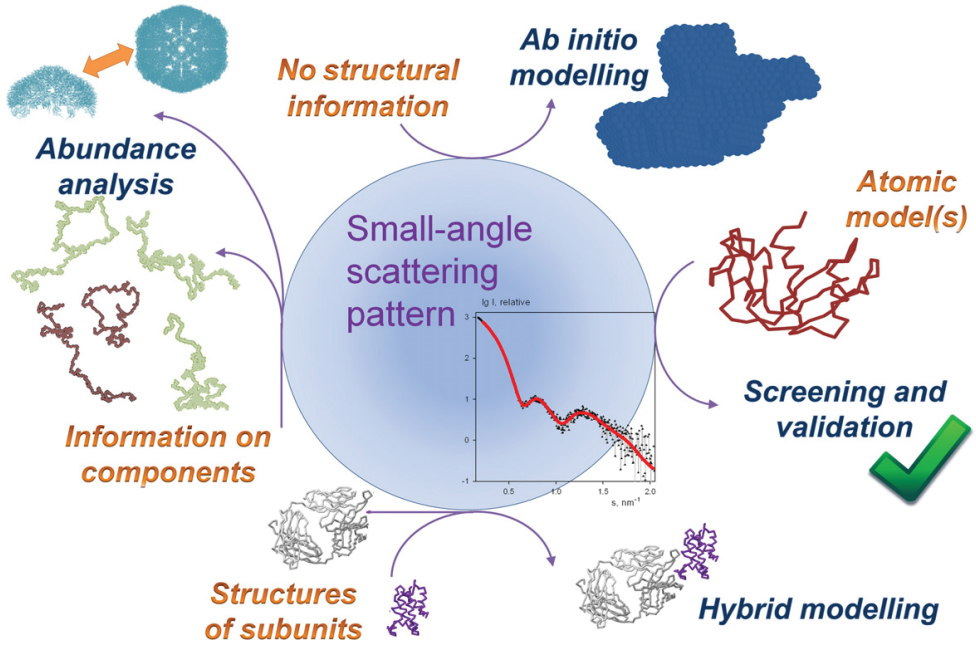

Figure 1.1: Overview of structural analysis options provided by SAXS. Figure was adapted from [18] with permission from Elsevier. Copyright ®2013 Elsevier Ltd..

The latter data are used to account for solvent scattering and to eliminate background scattering generated by the instrumental setup. The scattering contribution of the studied molecule depends on the molecular concentration (proportional to the number of particles focused by the beam) and the difference in electron density between the molecule and the solvent (typically denoted as contrast). The solute concentration has to be carefully balanced to enhance the signal-to-noise-ratio on the one hand, however, on the other hand, artifacts like aggregation can be caused by too high concentrations.

An important advantage of SAXS compared to other structure resolving techniques is that as a solution-based method it does not require crystallization or surface fixation of the molecules. Furthermore, compared to methods with size restrictions such as NMR, AFM or EM, SAXS allows for studying particles with molecular size ranging from a few kilo Daltons (kDa) up to large complexes of multiple giga Daltons (GDa) [19, 20]. Although SAXS only provides low resolution structural information, the obtained data such as electron density maps from *ab initio* modeling [15, 21–23], can be supplemented with high resolution substructures or scattering profiles to either refine theoretical predicted models [20, 24] or molecular dynamics simulations [25–27]. Moreover, the methods presented in the results section of this thesis demonstrate how small gold nanocrystals can be combined with small-angle scattering experiments to enhance the information content obtained from SAXS. The following chapters of this introduction provide a brief overview of the theoretical (Chapter 2) and experimental principles (Chapter 3) of small-angle X-ray scattering.

Chapter 2

Basic Theory of Small Angle Scattering

The following chapter outlines major theoretical concepts of small-angle scattering required to understand the experimental instrumentation in Chapter 3.

2.1 Elastic and Kinematic Scattering

2.1.1 Scattering by a Single Electron

Given by the wave-particle duality concept originating from quantum mechanics, X-ray photons cannot only be described in terms of particles but can also be considered waves. Furthermore, the energy E of particles is directly related to the wavelength λ by the term [28]:

$$E[\text{keV}] = \frac{hc}{\lambda} = \frac{12.398}{\lambda[\text{Å}]} \tag{2.1}$$

X-rays used in structural studies are typically in the 10 keV regime which corresponds to a wavelength λ of 0.10 nm or 1 Å [29–31]. Most applications require monochromatic X-ray waves which are characterized by the wavenumber $k = 2\pi/\lambda$. The mathematical description of an electromagnetic wave transversing along the z-axis with amplitude A is described by $A \cdot \sin(kz)$ in real coordinate space $\mathbb{R}$ or the complex exponential notation $A \cdot e^{ikz}$ in $\mathbb{C}$ [30].

To derive the scattering formalism for more complex materials, such as molecules, it is useful to start with the most elementary scattering object: a single, free electron. In the following paragraphs, the scattering length of an electron will be derived.

The first fundamental formalism of a scattering process is the differential scattering cross-section ($d\rho/d\Omega$) which is given by

$$\left(\frac{d\rho}{d\Omega}\right) = \frac{I_{sc}}{\Phi_0 \Delta\Omega} \tag{2.2}$$

using the strength of the incident beam Ψ_0 given by the number of photons passing through a unit area per second. The incident photons interact with the material and are scattered. A detector which is placed at a distance R to the sample records the number of scattered photons I_{sc} within an angle $\Delta\Omega$ deviating from the initial trajectory [30].

The next step is to apply the expression of an electromagnetic wave using the terms $\mathbf{E_{in}}$ for the incident electric field and $\mathbf{E_{rad}}$ to rewrite Eq. 2.2 to [30]

$$\left(\frac{d\rho}{d\Omega}\right) = \frac{|E_{rad}|^2 R^2}{|E_{in}|^2}. \tag{2.3}$$

Moreover, following the classical description of Rayleigh scattering, the electric field of the incoming photons forces electrons within the sample to vibrate [30]. Furthermore, the classical electron radius, also referred to as Thomson scattering length, is given by

$$r_o = \left(\frac{e^2}{4\pi\epsilon_0 m_e c^2}\right) = 2.82 \times 10^{-5}\text{Å} \tag{2.4}$$

with the elementary charge e, the speed of light c, the mass of an electron m_e and the electric constant ϵ_0 [28].

Using Maxwell's equation and the Thomson scattering length (Eq. 2.4) next, the magnitude of ratio between the radiated and the incident electric field of a single electron can be expressed as [30]

$$\frac{E_{rad}(R,t)}{E_{in}} = -\left(\frac{e^2}{4\pi\epsilon_0 m_e c^2}\right) \cdot \frac{e^{ikR}}{R} \cdot P. \tag{2.5}$$

In this equation, P is the polarization factor defined by the angle between the incident and the emitted wave Ψ as $P = 90° - \Psi$ and depends on the X-ray source used. A synchrotron, for example, can provide a linearly polarized beam in the horizontal plane [30].

The incoming wave can interact with a particle by the photoelectric effect, Compton, Raman or Rayleigh scatter. However, only Rayleigh scattering describes an elastic scattering process, which is the scattering phenomenon that is considered for small-angle scattering [30]. The scattered wave is denoted as $k' = |\mathbf{k}'|$ with wavenumber k' and wave vector $\mathbf{k}'$. Given that the kinetic energy is conserved in an elastic scattering process, the wavenumbers of both waves are equal, i.e. $k' = k$. The momentum transfer or wave vector transfer $\mathbf{q}$ is given by the wave vector of the incoming wave $\mathbf{k}$ and the scattered wave $\mathbf{k}'$ by [31]

$$\mathbf{q} = \mathbf{k}' - \mathbf{k} \tag{2.6}$$

and is usually expressed in units of Å^{-1} or nm^{-1}. In experiments, the scattering intensity $I(q)$ is recorded as a function of the scattering vector $\mathbf{q}$. Typically, the scattering momentum is given as a function of scattering angle 2θ and wavelength λ of the X-rays as $|\mathbf{q}| = q = 4\pi \sin(\theta)/\lambda$ [31].

2.1.2 Scattering from Multiple Electrons

Based on the elastic Thomson scattering process described for single electrons above, the next step is to expand the essential equations to more complex systems, such as to an arrangement of multiple scatterers. For example, atoms within a sample described by positions $\mathbf{r_i}$ (Fig. 2.1) can be considered as source of new spherical waves and given the correlated arrangement between the individual atoms the nett scattering signal is caused by interference between the emitted waves (see Fig. 2.1) [31]. Thus, the real space coordinates $\mathbf{r_i}$ have to be transfered into reciprocal space coordinates with scattering vectors $\mathbf{q_i}$ using Fourier transformation [30].

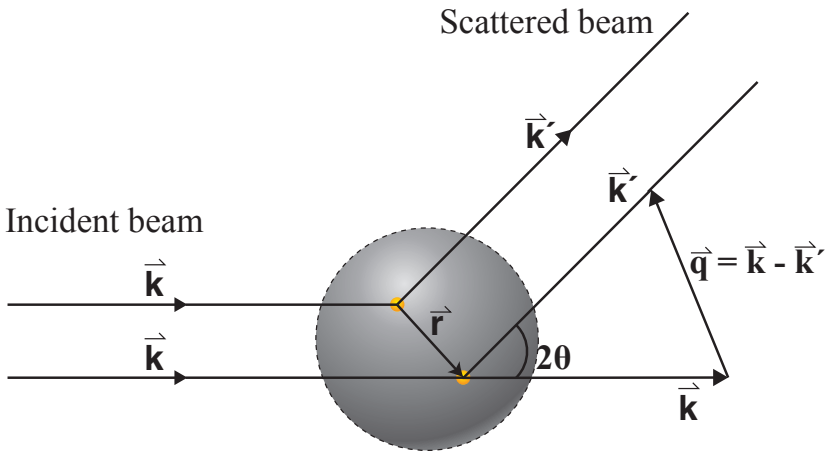

Figure 2.1: Schematic of a scattering process. The incident X-rays are elastically scattered off atoms (yellow circles) within a macromolecule (grey area). Due to the correlated positions of atoms with real space distance $||\mathbf{r}||$, the incident beam is coherently scattered. The momentum transfer stating the relation between the incident $\mathbf{k}$ and scattered wave vectors $\mathbf{k}'$ is given by $\mathbf{q} = \mathbf{k} - \mathbf{k}'$. Thus, scattered waves interfere because of the resulting phase shift. This figure was inspired by Fig. 1.2 of Ref. [31].

An important approximation used at this point (to allow for an easier description of systems with two or more electrons) is to consider scattering to be weak and therefore neglect multiple scattering effects. This is usually referred to as the kinematical approximation [32].

The simplest possible multi-particle system to look at is composed of two electrons which are considered to be structureless. They are also described as point scatterers [33]. Choosing the coordinate system in a way that one electron is placed in the origin the position of the other electron is given by the vector $\mathbf{r}$ [31]. Another approximation to do at this point is the far-field limit, which assumes that source and detector are sufficiently distanced from both electrons so that both the incident and scattered X-ray waves can be represented as plane waves [30]. The scattering vector for an elastic scattering process is given by $|\mathbf{q}| = (4\pi/\lambda) \cdot \sin(\theta)$ with the scattering angle 2θ as derived above [28]. Furthermore, the phase difference between the incoming wave specified by the wave vector $\mathbf{k}$ and the scattered wave $\mathbf{k}'$ is given by $\phi = (\mathbf{k} - \mathbf{k}') \cdot \mathbf{r}$ [31]. Using the definition of the momentum transfer $\mathbf{q}$ the equation can be rewritten to $\phi = \mathbf{q} \cdot \mathbf{r}$. Assuming that the polarization of the incident beam is perpendicular to the plane spanned by the incident wave vector $\mathbf{k}$ and the scattered wave vector $\mathbf{k}'$, the scattering amplitude can be expressed independently from the polarization factor (see Eq. 2.5) as the sum of both electromagnetic waves as [33]

$$A(\mathbf{q}) = -r_0(1 + e^{i\mathbf{qr}}) \tag{2.7}$$

and the intensity of the two-electron system being the product of the amplitude and its complex conjugate

$$I(\mathbf{q}) = A(\mathbf{q}) \cdot A(\mathbf{q})^* = 2r_0^2(1 + \cos(\mathbf{q} \cdot \mathbf{r})). \tag{2.8}$$

Using the definition of electromagnetic waves, as stated above, for a system with not only two,

but N electrons with the vector $\mathbf{r_i}$ pointing to the position of the i^{th} electron, this equation can be extended to a more general formalism and be written as [33]

$$A(q) = -r_0 \sum_{i=1}^{N} e^{(iq\cdot \mathbf{r_i})}. \tag{2.9}$$

It is important to point out that the derived description is only valid if the interaction of X-rays with the scattering material is weak [33]. Due to the weak interaction, the number of scattered photons is low and thus only a weak diffraction signal is produced. Nevertheless, modern X-ray sources allow to overcome this drawback providing highly intense X-ray beams.

Orientational Averaging

In order to calculate the scattering profile for a multi-electron system it is necessary to determine the geometrical orientation of the wave vector transfer $\mathbf{q}$ and the position vector $\mathbf{r}$. Using Eq. 2.8 and assuming a stationary system with a fixed angular relation between the two vectors $\mathbf{q}$ and $\mathbf{r}$ the scattering intensity for a two-electron system can be calculated (see Fig. 2.2a). To do that, a scattering amplitude f_1 was assigned to the electron positioned at the origin and the scattering amplitude f_2 to the second electron at position $\mathbf{r}$ [33]:

$$A(\mathbf{q}) = f_1 + f_2 e^{i\mathbf{q}\cdot\mathbf{r}}. \tag{2.10}$$

Thus, the intensity is given by [33]

$$I(\mathbf{q}) = f_1^2 + f_2^2 + f_1 f_2 e^{i\mathbf{q}\cdot\mathbf{r}} + f_1 f_2 e^{-i\mathbf{q}\cdot\mathbf{r}}. \tag{2.11}$$

However, for biological systems molecules adopt random orientations in solution. Hence another equation is necessary. The simple two-electron formalism needs to be expanded in a way that allows the position vector $\mathbf{r}$ to adopt randomly orientated directions [33].

Since the time required by X-rays to pass through the sample is much shorter than the typical timescale of molecular motions, an X-ray experiment records a series of snapshots of molecular orientations [33]. The integration time for experiments using in-house or synchrotron X-ray sources typically ranges from seconds to hours to obtain a sufficient signal-to-noise-ratio [34]. Thus the recorded snapshots are averaged at the detector. Nevertheless, the advance of high brilliant free electron X-ray lasers allows for studying molecular motions happing in the femtosecond timescale [7, 8].

To derive the formalism needed, the scattering amplitude f_1 is once more assigned to the first and the scattering amplitude f_2 to the second electron. The scattering intensity for a single snapshot is still given by Eq. 2.11. The distance between the two particles remains fixed. That fact implies that the length of $\mathbf{r}$ is constant. However, the direction of $\mathbf{r}$ is allowed to vary and to enclose randomly distributed angles with the momentum transfer vector $\mathbf{q}$. The measured intensity can be obtained by averaging over all possible orientations and assuming spherical symmetry (see Fig. 2.2b). The orientational averaged phase factor is given by [33]

$$\left\langle e^{i\mathbf{q}\cdot\mathbf{r}} \right\rangle_{orient.aver.} = \frac{\int e^{iqr\cdot\cos\theta} \sin\theta d\theta d\varphi}{\int \sin\theta d\theta d\varphi} \tag{2.12}$$

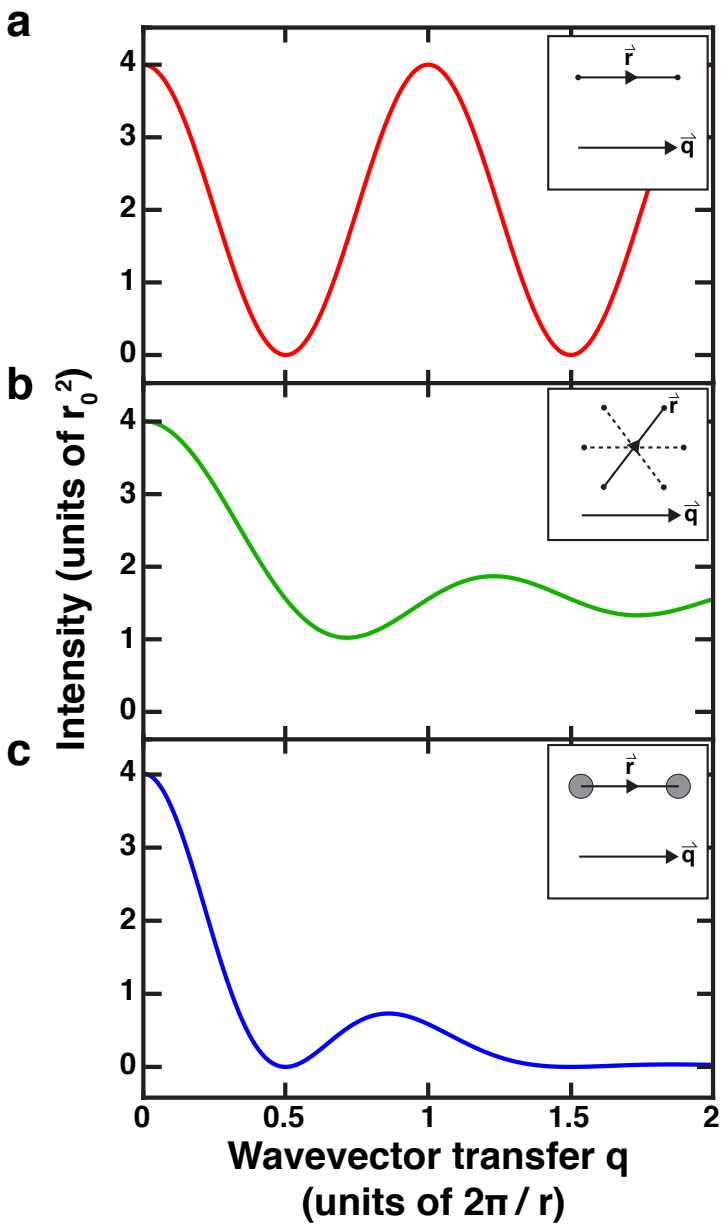

Figure 2.2: Diffraction pattern of multi-particle systems. (**a**) Diffraction pattern of two point-like electrons separated by a fixed vector **r** and parallel wave vector transfer **q**. The intensity was calculated using Eq. 2.11. (**b**) Diffraction pattern of two point-like particles separated by randomly orientated positions **r** with fixed length $r = ||\mathbf{r}||$. The intensity is given by Eq. 2.13. For both (a) and (b) the scattering amplitudes of the first particle f_1 and the second particle f_2 were set to the Thomson scattering length $f_1 = f_2 = -r_0$. (**c**) Diffraction pattern of two spherical scatterers separated by the distance $r = ||\mathbf{r}||$ and fixed position vector **r**. The scattering amplitude for this system was calculated with Eq. 2.19 derived in the next section and inserted in Eq. 2.11 to calculate the intensity profile. The radius of the spherical scatterer was set to $a/r = 0.25$. This figure was inspired by Fig. 4.2 of Ref. [33].

and can be substituted into Eq. 2.11

$$\left\langle I(\mathbf{q})\right\rangle_{orient.aver.} = f_1^2 + f_2^2 + 2f_1f_2\left\langle e^{i\mathbf{q}\cdot\mathbf{r}}\right\rangle_{orient.aver.}. \tag{2.13}$$

The integral in the enumerator of phase factor in 2.12 can be evaluated to [33]

$$\begin{aligned}\int e^{iqr\cdot\cos\theta}\sin\theta d\theta d\varphi &= 2\pi\int_0^{\pi} e^{iqr\cdot\cos\theta}\sin\theta d\theta = 2\pi\left(\frac{-1}{iq\cdot r}\right)\int_{iq\cdot r}^{-iq\cdot r} e^x dx \\ &= 4\pi\frac{\sin(q\cdot r)}{q\cdot r}\end{aligned} \tag{2.14}$$

while the denominator is

$$\int \sin\theta d\theta d\varphi = 2\pi\int_0^{\pi}\sin\theta d\theta = 4\pi. \tag{2.15}$$

Thus, the orientational averaged phase factor can be written as follows [31]:

$$\left\langle e^{i\mathbf{q}\cdot\mathbf{r}}\right\rangle_{orient.aver.} = \frac{\sin(q\cdot r)}{q\cdot r} \tag{2.16}$$

The diffraction pattern as a function of q for a two-particle system is shown in Fig. 2.2a in which the scattering particles were assumed to be with point-like scatterers with a scattering amplitude $f_1 = f_2 = r_0$ (r_0 is the Thomson scattering length) without a momentum transfer dependency. In case of biological macromolecules, the incident X-ray waves are scattered off the electron cloud surrounding atoms where the scattering amplitude f cannot longer be regarded to be independent from the momentum transfer (see Section 2.1.3). An example of two spherical scatterers with scattering amplitudes $f_1(\mathbf{q})$ and $f_2(\mathbf{q})$ is shown in Fig. 2.2c. The fact that electrons have a finite spatial size leads to a damping of the scattering intensity with increasing q [33]. The scattering amplitude of a solid spherical scatterer with electron density $\rho(r)$ and radius R is derived in the next section.

2.1.3 Scattering by a Single Atom

Having defined the scattering for point-like scatterers, the next step is to describe an atom with Z electrons. The mass of an electron m_e (9.109×10^{-31} kg) is much smaller as the mass of a proton m_p (1.673×10^{-27} kg) or a neutron m_n (1.675×10^{-27} kg) which together form the nuclei of atoms. This is why the vibration of electrons and their emitted intensities are high compared to the nucleons vibrations and emissions [30]. Following the classical description, the electron density of an atom is $\rho(r)$ [33]. The net charge within the cloud is given by the element number of the atom Z [28]. The charge within a volume element $d\mathbf{r}$ can therefore be computed via integrating over the charge density. To evaluate the scattering amplitude, the phase factor $e^{i\mathbf{q}\cdot\mathbf{r}}$ has to be included in the integral [30]. The calculated value $f^0(\mathbf{q})$ is usually denoted as atomic form factor and increases with atomic number Z. At the limit of zero momentum transfer $q\to0$, there is no phase shift between the scattered waves. Moreover, $f^0(\mathbf{q}=0)$ is equal to the number of electrons in the atom Z [35]. Since the wave vector transfer correlates with the wavelength of the incident X-ray waves by $2\pi/\lambda$, the wavelength of the radiation becomes

much smaller compared to the size of the atom at the limit of $\mathbf{q} \to \infty$ [28]. Thus, with an increasing momentum transfer q, more and more volume elements start to scatter out of phase and, at the limit of $q \to \infty$, the atomic form factor $f^0(\mathbf{q} \to \infty)$ reduces to 0 due to destructive interference [33].

$$f^0(\mathbf{q}) = \int \rho(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \tag{2.17}$$

In the quantum mechanics picture, the probability of finding an electron within an atom at any specific position around the nucleus is no longer given by a simple density function but rather by the wave-function $\psi_n(r)$ and the quantum number n. The wave-function of a one electron atom (similar to the hydrogen atom) in ground state is [33]

$$\psi_1(r) = \frac{1}{\sqrt{\pi a^3}} e^{-r/a}. \tag{2.18}$$

In this equation, the effective radius a of the electron distribution depends on the atomic number Z of the atom. The electron density is now given by $|\psi_1(r)|^2$. Thus, using Eq. 2.17 and the relation $\mathbf{q} \cdot \mathbf{r} = qr \cdot \cos\theta$ in a spherical coordinate system with the polar angle θ, the atomic form factor can be evaluated to [33]:

$$\begin{aligned}
f_1^0(\mathbf{q}) &= \frac{1}{\pi a^3} \int e^{-2r/a} \cdot e^{i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \\
&= \frac{1}{\pi a^3} \int_0^\infty 2\pi r^2 e^{-2r/a} \int_0^\pi e^{iqr\cdot\cos\theta} \sin\theta d\theta dr \\
&= \frac{1}{\pi a^3} \int_0^\infty 2\pi r^2 e^{-2r/a} \frac{1}{iqr} \left[e^{iqr} - e^{-iqr} \right] dr \\
&= \frac{1}{\pi a^3} \int_0^\infty 2\pi r^2 e^{-2r/a} \frac{2\sin(qr)}{qr} dr \\
&= \frac{4}{a^3} \frac{1}{q} Im \left[\int_0^\infty r e^{-r(2/a - iq)} dr \right] \\
&= \frac{1}{[1 + (qa/2)^2]^2}.
\end{aligned} \tag{2.19}$$

Moreover, the scattering factor for a solid spherical particle with radius R can be evaluated using the same approach. Under the assumption of a uniform charge distribution, the scattering

factor is given by [36]

$$
\begin{aligned}
f(\mathbf{q}) &= \int_0^\infty \rho(r) \cdot e^{i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} = \int_0^\infty 2\pi r^2 \rho(r) \int_0^\pi e^{iqr\cdot\cos\theta} \sin\theta d\theta dr \\
&= 4\pi \int_0^\infty \rho(r) \frac{\sin(q\cdot r)}{q\cdot r} r^2 dr = 4\pi \int_0^R \frac{\sin(q\cdot r)}{q\cdot r} r^2 dr \\
&= \frac{4\pi}{q} \int_0^R \sin(q\cdot r) r dr = \\
&= \frac{4\pi}{q} \left(-\frac{R\cos(q\cdot R)}{q} + \left[\frac{\sin(q\cdot r)}{q} \right]_0^R \right) \\
&= \frac{4\pi}{q} \left(-\frac{R\cos(q\cdot R)}{q} + \frac{\sin(q\cdot r)}{q^2} \right) \\
&= \frac{4}{3}\pi R^3 \frac{3[\sin(q\cdot R) - q\cdot R\cdot\cos(q\cdot R)]}{(q\cdot R)^3}
\end{aligned}
\tag{2.20}
$$

with a charge density of:

$$
\rho(r) = \begin{cases} 1 & \text{if } 0 \le r \le R \\ 0 & \text{if } r > R \end{cases} \tag{2.21}
$$

Nevertheless, the evaluated expression for the atomic scattering factor only holds up in cases were no absorption or other energy dependent processes can occur. As will be explained in Section 2.2.2, the scattering factor of an atom has to be extended by a dispersion correction term $f' + if''$ in cases as such, taking the binding energies of electrons into account [28].

2.1.4 Scattering from a Molecule

After having derived the scattering description for point-like particles and single atoms, the next step is a group of atoms organized to a molecule. The scattering amplitude for a molecule with N atoms and individual atomic scattering factors f_i is given by [30]

$$
F_{mol}(\mathbf{q}) = \sum_{i=1}^{N} f_i(\mathbf{q}) \cdot e^{i\mathbf{q}\cdot\mathbf{r_i}}. \tag{2.22}
$$

Using the expression derived for the orientational averaged phase factor 2.16, the scattering

profile of a molecule is [33]

$$\begin{aligned}\left\langle \left| \sum_{i=1}^{N} f_i e^{i\mathbf{q}\cdot\mathbf{r_i}} \right|^2 \right\rangle_{orient.aver.} &= \left|f_1\right|^2 + \left|f_2\right|^2 + \ldots + \left|f_N\right|^2 \\ &+2f_1f_2\frac{\sin(q\cdot r_{1,2})}{q\cdot r_{1,2}} + 2f_1f_3\frac{\sin(q\cdot r_{1,3})}{q\cdot r_{1,3}} + \ldots + 2f_1f_N\frac{\sin(q\cdot r_{1,N})}{q\cdot r_{1,N}} \\ &+2f_2f_3\frac{\sin(q\cdot r_{2,3})}{q\cdot r_{2,3}} + \ldots + 2f_2f_N\frac{\sin(q\cdot r_{2,N})}{q\cdot r_{2,N}} \\ &+ \ldots + 2f_{N-1}f_N\frac{\sin(q\cdot r_{N-1,N})}{q\cdot r_{N-1,N}}.\end{aligned} \tag{2.23}$$

This equation was first published by Peter Debye in 1915 [37] and is commonly written as

$$I(q) = \sum_{i=1}^{N}\sum_{j=1}^{N} f_i(q)f_j(q)\frac{\sin(qr_{(i,j)})}{qr_{i,j}}. \tag{2.24}$$

Here, $r_{i,j} = |\mathbf{r_i} - \mathbf{r_j}|$ corresponds to the interparticle distances between the different scatterers.

Another valid approach to calculate the amplitude of a particle, discovered by Stuhrmann [38, 39], is to use spherical harmonic expansions

$$F_P(\mathbf{q}) = \sum_{i=1}^{N} f_i(q)\cdot e^{i\mathbf{q}\cdot\mathbf{r}_i} = 4\pi\sum_{l=0}^{L}\sum_{m=-l}^{l} i^l Y_{lm}(\Omega)\sum_{i=1}^{N} f_i(q)j_l(qr_i)Y^*_{lm}(\omega_i) \tag{2.25}$$

with spherical Bessel functions $j_l(qr)$ of order l, their corresponding spherical harmonics $Y_{lm}(\Omega)$ of order l and m, the truncation value L and $f_i(q)$, the atomic form factors of the individual elements. Using Eq. 2.25, the scattering intensity is obtained by [31]:

$$I(q) = \langle F_P(\mathbf{q})\cdot F_P^*(\mathbf{q})\rangle_{orient.aver.} \tag{2.26}$$

Solution Scattering

Given that SAXS is a solution-based method, the electron density of the solutions used cannot be neglected. In addition, the diffraction pattern originates from deviations in electron density between the molecule and the surrounding buffer [31]. Furthermore, assuming a monodisperse system with average electron density ρ_{av} and neglecting short range interaction for non-crystalline samples, the scattering intensity for solution based small-angle scattering is [33]

$$I_{SAS}(q) = f^2\sum_i\int_V \rho_{av}\cdot e^{i\mathbf{q}\cdot(\mathbf{r_i}-\mathbf{r_j})}dV_j. \tag{2.27}$$

In this equation, dV_j is the volume element located at position $\mathbf{r_j} - \mathbf{r_i}$ with respect to the atom at position $\mathbf{r_i}$ [33]. Experimentally, separate measurements of the scattering signal from the background solution and the solvent are required. Assuming a solution with constant scattering length density ρ_{solv}, the difference in electron density between the solution and the

solvent ρ_{av} results in the excess scattering length density $\Delta\rho(\mathbf{r}) = \rho_{av} - \rho_{solv}$ also denoted to as contrast [31]. Using this , the scattering intensity yields [40]

$$\begin{aligned} I(\mathbf{q}) &= \left| \int_V \Delta\rho(\mathbf{r}) \cdot e^{i\mathbf{q}\cdot\mathbf{r}} dV \right|^2 = (\rho_{av} - \rho_{solv})^2 \left| \int_{V_P} e^{i\mathbf{q}\cdot\mathbf{r}} dV_P \right|^2 \\ &= \Delta\rho^2 V_p^2 |F(\mathbf{q})|^2 \end{aligned} \tag{2.28}$$

introducing the so-called form factor

$$F(\mathbf{q}) = \frac{1}{V_P} \int_{V_P} e^{i\mathbf{q}\cdot\mathbf{r}} dV_P \tag{2.29}$$

which depends on the shape and the size of the molecule.

Contrast and Resolution

While nucleic acids have a high average electron density of ~0.55 $e^-/\text{Å}^3$ [41] due to the phosphorus atoms incorporated in the backbone, the difference in average electron density of proteins (~0.42 $e^-/\text{Å}^3$ [31]) compared to pure water (~0.33 $e^-/\text{Å}^3$ [13]) is smaller which results in a weaker scattering signal for equally sized molecules. In comparison, heavy atoms, such as gold, have an electron density (~19.3 $e^-/\text{Å}^3$) over 10-fold higher [41]. However, usually monovalent and divalent ions are added to the solution to probe the molecule under physiological conditions or titration series ranging from low to high salt are used to probe electrostatic effects and related conformational changes [20]. The addition of salts or other required molecules, as for example sucrose, reduces the difference in electron density further and thus lowers the contrast [42]. For example, the addition of 1 M NaCl reduces the electron contrast from $0.854 \times 10^{-1} e^-/\text{Å}^3$ to $0.701 \times 10^{-1} e^-/\text{Å}^3$ in comparison to pure water [31]. The electron density of the solute can almost be matched using a high concentration of sucrose in the buffer [42]. However, salts or smaller molecules need to be added carefully as this procedure can perturb the molecule and induce conformational changes.

As indicated in Section 2.1.2, SAXS data are recorded in reciprocal space and transferred into real space via Fourier Transformation. Thus, a scattering signal recorded at small q-values corresponds to large distances in real space and large q-values correspond to short distances in real space [31]. Therefore the best possible resolution of a SAXS experiment is given by the highest detectable q-value q_{max}, which corresponds to information of the real space object with $\Delta = 2\pi/q$ [32]. While the diffraction pattern can be recorded to high q-values in crystallography providing details on the atomic scale, only low-resolution information can be obtained for SAXS studies due to the damping effect of spherical averaging and the resulting shorter q-range [31]. It should be noted that the scattering profiles obtained from small-angle scattering cannot be used to build unique high-resolution structure models but can provide precise information about size and shape of macromolecules [43].

2.2 Absorption and Anomalous X-ray Scattering

2.2.1 Absorption

Operating on the atomic scale, one also has to consider quantum mechanic interactions. Besides being elastically scattered off the electrons within the targeted sample, the incoming radiation can also be absorbed by the matter reducing the intensity of the transmitted beam [44]. This process is called photoelectric absorption. One important parameter is the material specific absorption cross-section, also called linear absorption coefficient μ, which indicates the capability of matter to absorb incoming X-rays [28].

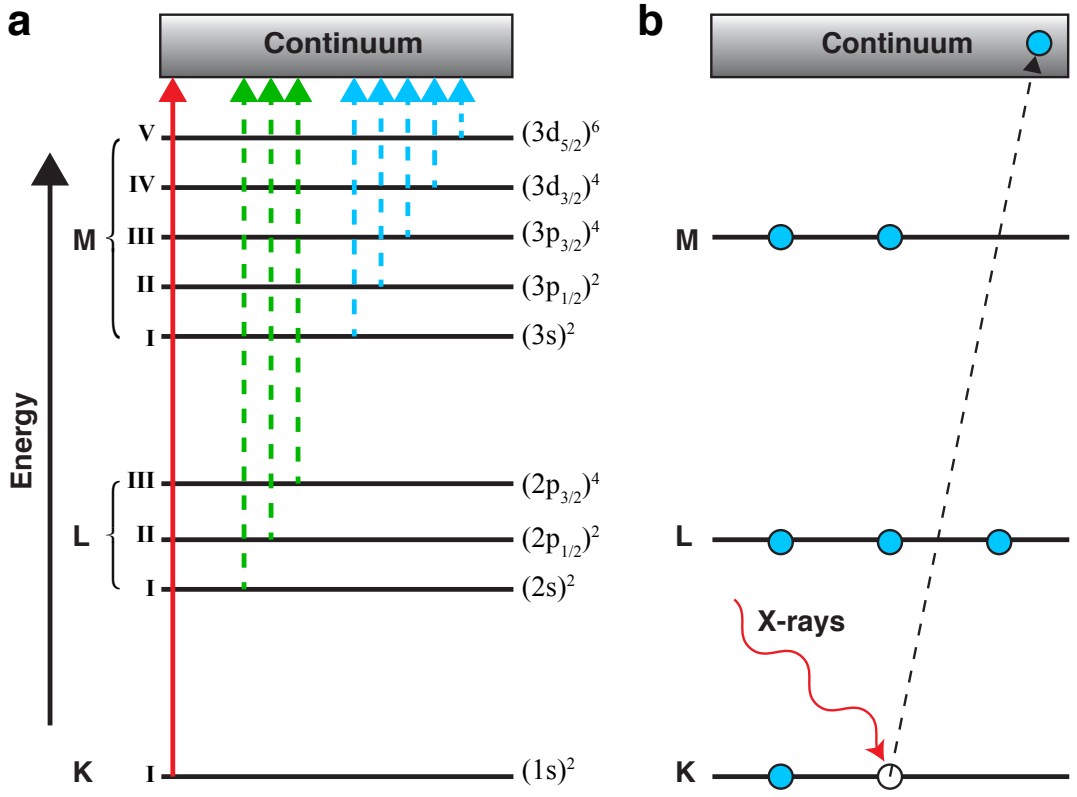

Figure 2.3: Schematic of energy levels and X-ray absorption. (**a**) Nomenclature for electron levels used to label the absorption edges. The arrows indicate possible transitions from the individual bound states to the continuum of free states. Additionally, states within the K, L and M shell are labeled using the principal, orbital angular momentum and the total angular momentum quantum numbers (n, l and j respectively) as (nl_j^{2j+1}). This figure was inspired by Fig. 7.2 form Ref. [44]. (**a**) An incident X-ray photon is absorbed by an electron within the K shell and is removed from the atom to the continuum. The emerging hole can be filled by an electron from a higher shell. This figure was inspired by Fig. 1.11 from Ref. [30].

The relation between the attenuation of an X-ray beam dI passing a sample with thickness dz and the intensity $I(z)$ at depth z from the entry point is given by [31]

$$-dI = I(z)\mu dz. \tag{2.30}$$

Eq. 2.30 can be rewritten to the following differential equation [31]

$$\frac{dI}{I(z)} = -\mu dz. \tag{2.31}$$

The equation can be solved choosing the approach that the intensity at $z = 0$ corresponds to the intensity of the incoming X-ray beam I_0. The solution provides an expression in a way that the transmitted beam can be expressed as a function of the absorption coefficient μ, the thickness of the sample z and the intensity of the incident beam $I(z = 0) = I_0$ [31]:

$$I(E) = I_0(E) \cdot e^{-\mu(E)z} \tag{2.32}$$

The absorption coefficient μ is related to the energy-dependent absorption cross-section per atom $\rho_{atom}(E)$ by following equation [28]:

$$\mu(E) = \left(\frac{\rho_{mass} N_A}{M}\right) \rho_{atom}(E). \tag{2.33}$$

Here, N_A is the Avogadro's number, ρ_{mass} the mass density and M the molar mass of the material. Eq. 2.32 can be rearranged to obtain an expression for the transmission T which indicates the ratio between the transmitted beam intensity I and the incident beam intensity I_0 [44]:

$$T = \frac{I}{I_0} = e^{-\mu(E)z} \tag{2.34}$$

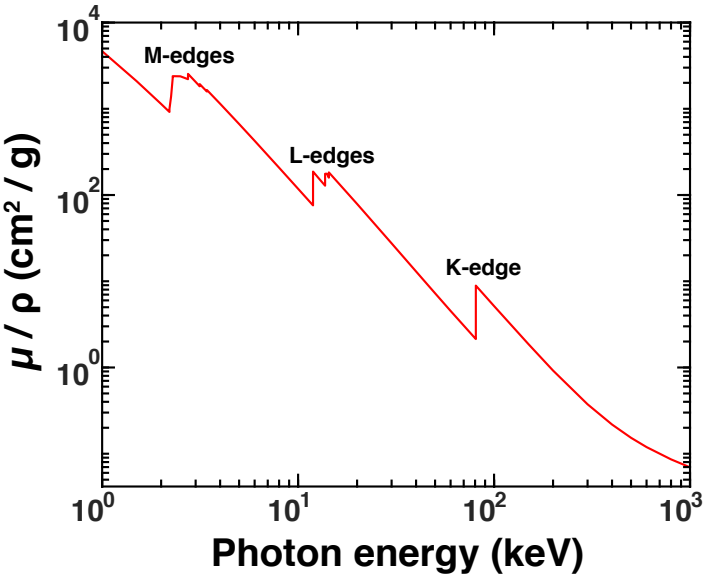

Figure 2.4: X-ray mass attenuation coefficient μ/ρ_{mass} as a function of photon energy for gold. Atomic absorption edges K, L and M are labeled. Data obtained from Ref. [45]

Moving from the classical picture of an atom with electron cloud density $\rho_{atom}(\mathbf{r})$ to the quantum mechanical description, electrons are no longer homogeneously distributed around the nuclei of an atom, but rather have precise energy levels being bound to the atomic core

in electron shells [44]. The electron shell with the highest binding energy is referred to as the K shell, followed by the less tightly bound L, M, N, O, P and Q shells (Figs. 2.3a, 2.4) [28]. Considering the binding energy of electrons within those shells, it requires a certain energy to lift an electron from a lower shell to a higher unoccupied position, or to entirely remove the electron from the atom to the continuum of free states (Figs. 2.3b, 2.4) [44]. Moreover, the energy levels described by the principal quantum number n is further split into sub-levels by spin-orbit coupling. The convention to denote these states makes use of the angular momentum quantum number l and the number of orbitals $2j_l + 1$ within a sub-shell given by the magnetic quantum number j_l [28]. Usually, the angular momentum quantum number l is written in the following letter code:

l	0	1	2	3	...
Letter	s	p	d	f	...

However, it is more convenient to introduce the new term of the mass absorption coefficient μ/ρ_{mass} (Fig. 2.4) for absorption calculations since it is constant for an individual element and additive for a mixture of multiple components [46]. While weighting each component by the factor w_i for the fraction in the composite, the mass absorption coefficient of a mixture can immediately be written as [44]:

$$\left(\frac{\mu}{\rho_{mass}}\right)_{mixture} = \sum_i w_i \left(\frac{\mu}{\rho_{mass}}\right)_i \tag{2.35}$$

2.2.2 Anomalous X-ray scattering

The atomic form factor was introduced in Section 2.1 for elastic scattering in terms of classical Thomson scattering. Nevertheless, as outlined in Section 2.2.1, the scattering of X-rays cannot be described without considering energy exchange, treating the scatterers as extended distributions of free electrons. In the model of bound electrons introduced in Section 2.2.1, the electrons occupy defined energy levels. Thus, the electrons cannot respond to the electric field of incident X-rays freely [47]. But an additional energy-dependent term needs to be included to describe the atomic form factor $f^0(\mathbf{q})$ of an atom [30]. The response of bound electrons to a driving field is characterized by a classical damped harmonic oscillator with resonant frequency ω_{res} and a damping constant γ [28].

Hence, the atomic form factor $f^0(\mathbf{q})$ for a free electron has to be reduced by an additional term typically denoted as f'. For energies much greater than the binding energy, the electrons can be assumed to be free and thus $f' = 0$ [47]. However, for energies close to energy levels given by the electron shell, f' displays resonant behavior [30]. These energy levels are known as the atomic absorption edges of a material [44]. In addition to the real part of the scattering length f', a phase lag with respect to the driving field alters the atomic scattering factor close to an absorption edge by a term if'' [31]. The phase lag causes negative interference between the incoming and the emitted radiation and is therefore imaginary. Including the dispersion correction terms f' and f'' the atomic form factor f can altogether be written as [48]

$$f(\mathbf{q}, \omega) = f^0(\mathbf{q}) + f'(\omega) + if''(\omega). \tag{2.36}$$

Both f' and f'' are maximized when the X-ray energy is equal to the energy of an absorption edge of an atom [49]. They are referred to as resonant scattering terms. The absorption effect can be described by the dispersion corrections of the total scattering cross-section [49]. While higher order orbitals are sparsely populated for light atoms, the dispersion correction is dominated by electrons in the K shell [47]. These orbitals are more densely occupied for heavier elements (i.e. gold) and thus expand the resonant effect to higher order shells (L or M shell) [45]. While the classical Thomson scattering term $f^0(\mathbf{q})$ does not depend on the X-ray energy but only on the momentum transfer $\mathbf{q}$ and the atomic number Z, the scattering vector dependency can be neglected for the dispersion correction terms f' and f'' due to the spatial confinement of electron orbitals [49]. Furthermore, the imaginary component is related to the photo absorption cross-section σ_{atom} via the following expression [50]:

$$f''(0) = \frac{\sigma_{atom}}{2r_0\lambda} \tag{2.37}$$

with the classical Thomson scattering length r_0 and the X-ray wavelength λ.

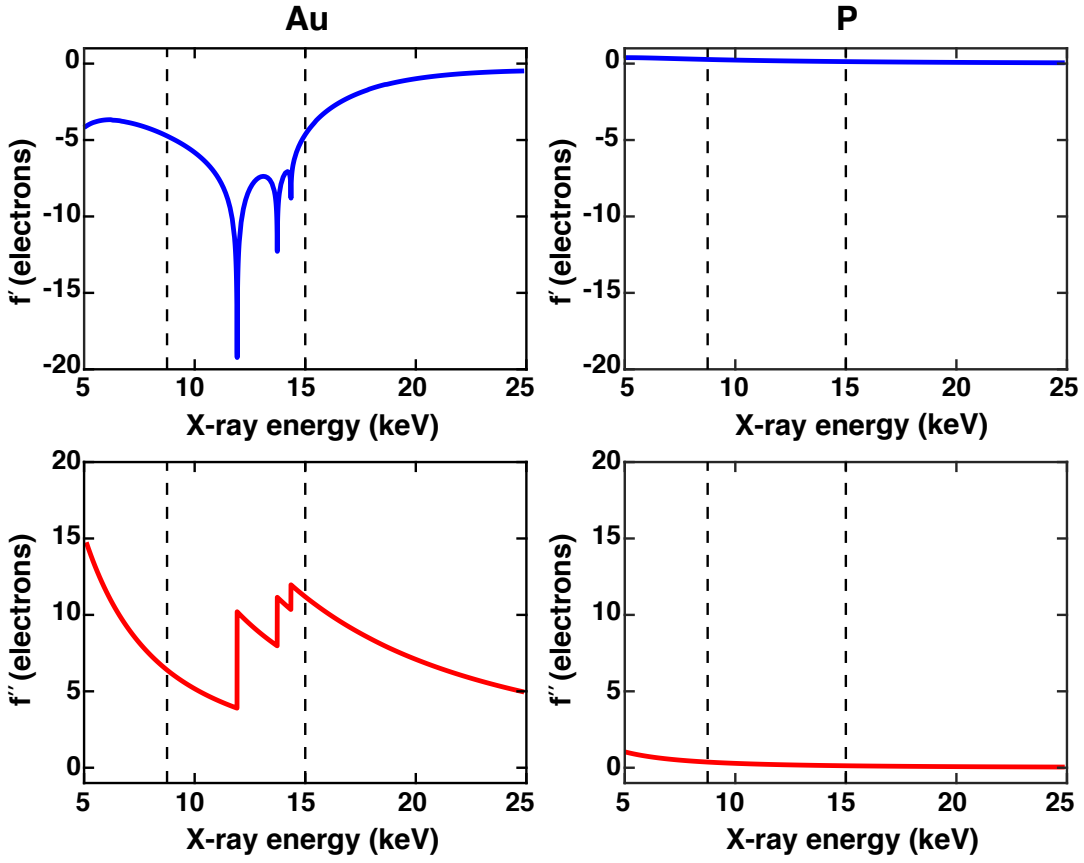

Figure 2.5: Real and imaginary part of the dispersion corrections. The energy dependency of the real and the imaginary part of the dispersion corrections for gold (Au) and phosphorus (P). The L-I, L-II and L-III absorption edges with significant changes in f' and f'' are nicely visible for gold while there is almost no dispersion correction for phosphorus. The dashed lines indicate the typical X-ray energy range for SAXS experiments from 9 keV up to 15 keV. Data obtained from Ref. [51]

Damped Single Oscillator Model

As stated above, the simplest model to describe an electron bound in an atom is by using a charged damped oscillator. The equation for an oscillator with the driving field $\mathbf{E_{in}} = \hat{\mathbf{x}} \cdot E_0 e^{-i\omega t}$, linearly polarized along the x-axis with amplitude E_0 and frequency ω, is [49]

$$\ddot{x} + \gamma \dot{x} + \omega_{res}^2 x = -\left(\frac{eE_0}{m}\right) e^{-i\omega t}. \tag{2.38}$$

Here, ω_{res} is the resonant frequency, γ is a damping constant, m is the mass and e is the elementary charge.
The damping of the oscillator represents the loss of energy due to re-radiation of excited electrons jumping to a lower energy level or back to the ground state. Using the trivial solution for a differential equation, $x(t) = x_0 e^{-i\omega t}$, the amplitude of the charged oscillator can be gained [49]:

$$x_0 = -\left(\frac{eE_0}{m}\right) \frac{1}{(\omega_{res}^2 - \omega^2 - i\omega\gamma)}. \tag{2.39}$$

Moreover, the strength of an electric field at distance R and time t depends on the acceleration $\ddot{x}(t - R/c)$ caused by the electrons at the time $t' = t - R/c$ [49]:

$$E_{rad}(R, t) = \left(\frac{e}{4\pi\epsilon_0 R c^2}\right) \ddot{x}(t - R/c) \tag{2.40}$$

Inserting the acceleration $\ddot{x}$ and the amplitude x_0 of the oscillator (obtained from Eq. 2.39) leads to [49]

$$\frac{E_{rad}(R, t)}{E_{in}} = -r_0 \frac{\omega^2}{(\omega_{res}^2 - \omega^2 - i\omega\gamma)} \left(\frac{e^{ikR}}{R}\right). \tag{2.41}$$

Since the atomic scattering length f is defined to be the amplitude of an outgoing spherical wave e^{ikR}/R in units of the Thomson scattering length $-r_0$, it can immediately be extracted and written as [52]:

$$f_c = \frac{\omega^2}{(\omega_{res}^2 - \omega^2 - i\omega\gamma)} \tag{2.42}$$

Finally, the real part of the dispersion correction f' is given by [49]

$$f_c' = \frac{\omega_{res}^2(\omega^2 - \omega_{res}^2)}{(\omega^2 - \omega_{res}^2)^2 + (\omega\gamma)^2} \tag{2.43}$$

and the imaginary part can be written as

$$f_c'' = -\frac{\omega_{res}^2 \omega\gamma}{(\omega^2 - \omega_{res}^2)^2 + (\omega\gamma)^2}. \tag{2.44}$$

Typically, the damping factor γ is much smaller than the the resonance frequency ω_{res}. Moreover, assuming X-ray energies much higher than the energy level of an absorption edge $\omega \gg \omega_{res}$, the dispersion correction reduces to $f_c = 1$ which corresponds to a free electron [49].

Damped Multiple Oscillator Model

While the model of a single damped oscillator is only able to yield a single energy level corresponding to a single excited state, there is a continuum of unoccupied free states as depicted in Section 2.2.1. Thus, a more complex model is required. Each energy level is associated with its distinct resonance frequency ω_{res} and every transition in between the states has a certain probability [49]. Thus, the generalized expression for the photo absorption cross-section of an ensemble of oscillators providing a continuum of excited states is

$$\sigma_a(\omega) = 2\pi^2 r_0 c \sum_{res} g(\omega_{res})\delta(\omega - \omega_{res}). \tag{2.45}$$

In this equation, $g(\omega_{res})$ is the weight of each transition [52]. Furthermore, the real part of the dispersion correction f' has to be considered as the superposition of an assembly of oscillators, too [49]:

$$f'(\omega) = \sum_{res} g(\omega_{res}) f'_c(\omega_{res}, \omega) \tag{2.46}$$

Kramers-Kronig Relation

Instead of using purely theoretical values of the dispersion corrections, more accurate values can commonly be achieved using the following approach: *i*) The imaginary part f'' is determined via extended X-ray absorption fine structure (EXAFS) measurements and *ii*) the real part f' is calculated from f'' using the Kramers-Kronig relation [53].

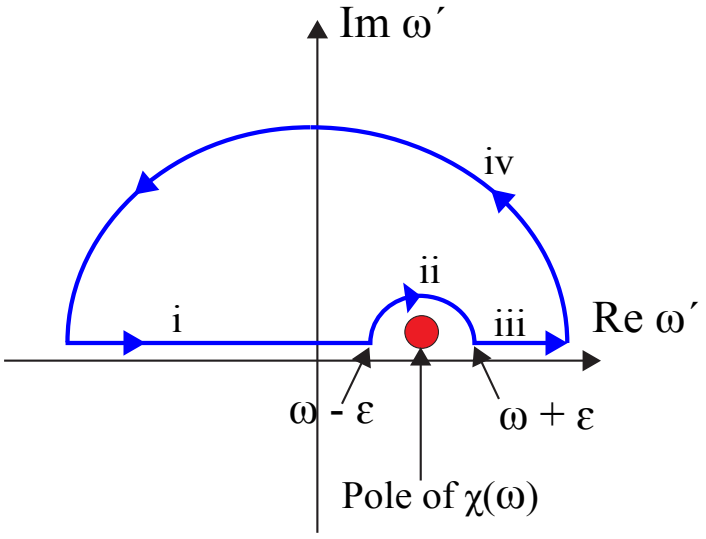

Figure 2.6: Integral contour for the Kramers-Kronig relation. The contour integral traces a semi-circle around a pole [54].

The Kramers-Kronig relation is based on Cauchy's theorem for complex integration linking the real and imaginary parts of any analytical function F(z) in the complex plane. If the theorem

is applied to the sum of the dispersion correction terms $\chi(z) = \omega_{res}^2/(z^2 - \omega_{res}^2 + iz\gamma)$, Cauchy's theorem yields

$$i\pi \left(f_c'(\omega) + if_c''(\omega)\right) = \mathcal{P} \int_{-\infty}^{\infty} \frac{f_c'(\omega') + if_c''(\omega')}{\omega' - \omega} d\omega'. \tag{2.47}$$

In this equation, $\mathcal{P}$ denotes Cauchy's principal value [49]. Eq. 2.47 can be rearranged so that the real part of the dispersion correction can be written as the Kramers-Kronig relations. The real part is expressed as [55]

$$f'(\omega) = \frac{1}{\pi}\mathcal{P} \int_{-\infty}^{\infty} \frac{f''(\omega')}{(\omega' - \omega)} d\omega' = \frac{2}{\pi}\mathcal{P} \int_{0}^{\infty} \frac{\omega' f''(\omega')}{(\omega'^2 - \omega^2)} d\omega' \tag{2.48}$$

and the imaginary part is given by:

$$f''(\omega) = -\frac{1}{\pi}\mathcal{P} \int_{-\infty}^{\infty} \frac{f'(\omega')}{(\omega' - \omega)} d\omega' = -\frac{2\omega}{\pi}\mathcal{P} \int_{0}^{\infty} \frac{f'(\omega')}{(\omega'^2 - \omega^2)} d\omega'. \tag{2.49}$$

The initial integrals in Eqs. 2.48 and 2.49 are decomposed into the four segments of the contour (see Fig. 2.6) in the following way: *i*) tracing the real axis from $-\infty$ to $\omega - \epsilon$; *ii*) a half circle around the pole; *iii*) again along the real axis from $\omega + \epsilon$ to ∞; and *iv*) back to the origin of *i*) along the large semi-circle. The integrals can be further collapsed using the fact that $f'(\omega')$ is an even function (Eq. 2.43) and $f''(\omega')$ an odd function (Eq. 2.44) [49].

2.3 Structural Parameters in SAXS

After having derived the key formalisms for elastic small-angle scattering this section provides an overview of important structural parameters in SAXS such as the molecular weight (MW), the radius of gyration (R_g), and the maximum intermolecular distance D_{max}. These parameters can be useful to assess the structural information, the conformational transition between multiple states or the formation of aggregates in solution.

2.3.1 Pair-Distance Distribution Function

While SAXS profiles are recorded in reciprocal-space, it is often easier to work and interpret real-space data. However, as shown in Section 2.1, the Fourier transformation involves an integral ranging from 0 to infinity momentum transfer, a range that is experimentally not feasible to measure. One important concept to describe structural parameters is the pair-distance distribution function which will be derived in this section.

First, the definition of the cross-section as well as Eqs. 2.22 and 2.28 are used [31]:

$$I_{total}(q) = \frac{\langle A(\mathbf{q}) \cdot A^*(\mathbf{q})\rangle}{V} = \frac{1}{V}\left\langle \int_V d\mathbf{r} \int_V d\mathbf{r}' \Delta\rho(\mathbf{r})\Delta\rho(\mathbf{r}') \cdot e^{-i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')} \right\rangle \tag{2.50}$$

Again, temporal and spherical averaging are applied while the integral is performed over the whole exposed sample volume V. Using the definition of the form factor $F(q)$ (Eq. 2.29) it

follows that [31]:

$$I_{total}(q) = \frac{N}{V} \underbrace{\left\langle \int_{V_P} d\mathbf{r} \int_{V_P} d\mathbf{r}' \Delta\rho(\mathbf{r})\Delta\rho(\mathbf{r}') \cdot e^{-i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')} \right\rangle}_{\text{form factor } F(q)} \cdot \underbrace{\left\langle \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} e^{-i\mathbf{q}\cdot(\mathbf{r_i}-\mathbf{r_j})} \right\rangle}_{\text{structure factor S(q)}} \tag{2.51}$$

The second factor is known as the structure factor $S(q)$ which is used to describe inter-particle interactions [33].

Introducing the pair-correlation function g(**r**) between the centers of mass of the particles the structure factor can be rewritten to [31]:

$$\begin{aligned} S(q) &= 1 + \frac{1}{N} \left\langle \frac{1}{N} \sum_{i=1}^{N} \sum_{j\neq i}^{N} e^{-i\mathbf{q}\cdot(\mathbf{r_i}-\mathbf{r_j})} \right\rangle \\ &= 1 + \frac{N-1}{V} \int g(\mathbf{r}) \cdot e^{-\mathbf{qr}} d\mathbf{r} \end{aligned} \tag{2.52}$$

Under the assumption of a dilute sample without any inter-particle interactions [31], the structure factor can be approximated to $S(q) \approx 1$ and thus the total scattering intensity $I_{total}(q)$ to be proportional to the first factor in brackets $F(q)$ in Eq. 2.51.

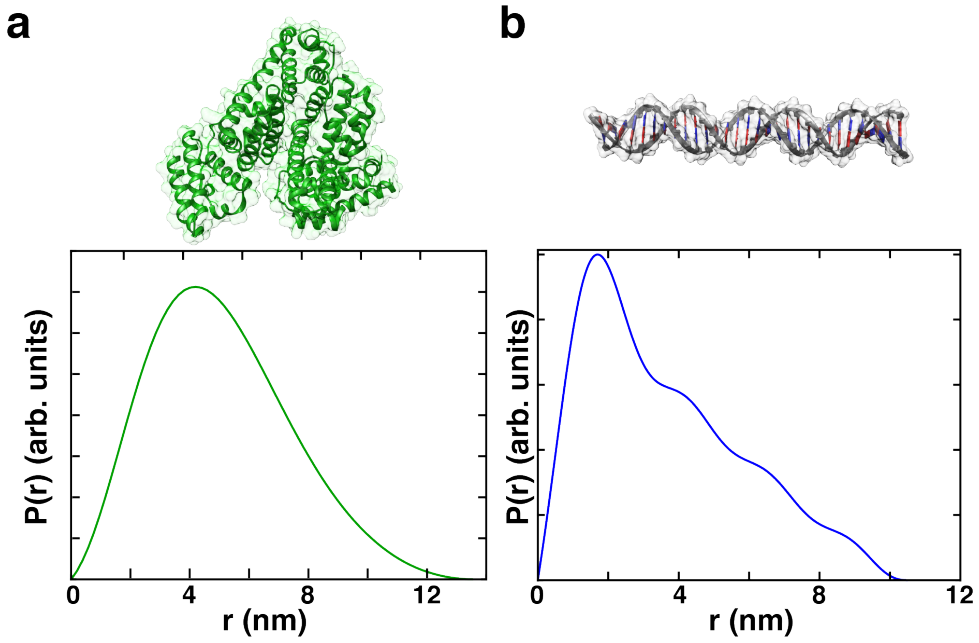

Figure 2.7: Pair-distance distribution function for different shapes. (**a**) Distance distribution function for the protein bovine serum albumin (BSA). The intensity profile was calculated from PDB code 3V03 (Ref. [56, 57]). (**b**) $P(r)$ function for a cylindrical shaped 30 base pair (bp) DNA macromolecule. The PDB file was obtained from Ref. [58]. Theoretical scattering profiles were calculated for both PDB files via Crysol [59] and distance distribution functions were obtained from Primus [60]. The difference in $P(r)$ for a nearly globular shaped particle compared to a rod-like shape is nicely visible comparing (a) and (b). Molecular graphics were prepared using *UCSF Chimera* [61].

Second, following Debye and Bueche [62], the correlation function of the particle $\gamma(\mathbf{r})$ is defined as:

$$\gamma_P(\mathbf{r}) = \left\langle \int_P \Delta\rho(\mathbf{r}')\Delta(\mathbf{r}' - \mathbf{r})d\mathbf{r}' \right\rangle \tag{2.53}$$

Substituting the correlation function Eq. 2.53 into the the scattering intensity Eq. 2.51, it follows that [31]:

$$I(q) = \left\langle \int_{V_P} \gamma_P(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \right\rangle = \int_{V_P} \left(\frac{1}{4\pi} \int_\Omega \gamma_P(\mathbf{r}) d\Omega \right) \cdot e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \tag{2.54}$$

Finally, solving the integral yields

$$I(q) = 4\pi \int_0^{D_{max}} P(r) \frac{sin(qr)}{qr} dr. \tag{2.55}$$

Here, the pair-distance distribution function $P(r) = r^2\gamma(\mathbf{r})$ [63] represents the number of distances between pairs of points within the distance interval from r to $r + dr$ inside the object and the maximum particle distance D_{max} [64]. Moreover, the distances are weighted by the product of the scattering length densities of the two points gained by the contrast in electron density to the solvent [31]. Fig. 2.7 shows the distance distribution functions for the globular protein bovine serum albumin (BSA) and a 30 base pair (bp) DNA duplex.

2.3.2 Guinier Analysis and Radius of Gyration

Guinier analysis refers to the long-wavelength limit $qr \to 0$, where the power series expansion of the function $\sin(x)/x$ obtained for orientation verged diffraction (see 2.55) pattern yields [65]:

$$\frac{\sin(qr)}{qr} = 1 - \frac{(qr)^2}{3!} + \frac{(qr)^4}{5!} - O((qr)^6). \tag{2.56}$$

A further simplification of Eq. 2.56 is achieved by a number of steps (see Refs. [17, 66] for more details) and results in:

$$I(q) = I(0) \left(1 - \frac{q^2 r^2}{3} + O(q^4) \right), \tag{2.57}$$

which can be transformed into the Guinier formula [31]

$$I(q) = I(0) \cdot e^{-q^2 r^2/3} = I(0) \cdot e^{-q^2 R_g^2/3}. \tag{2.58}$$

In this last equation, R_g is referred to as *radius of gyration*. It is defined as the sum of root-mean-squared distances of all elemental scattering volumes from their center of mass weighted by their scattering densities [67] and can be obtained by the Guinier approximation [17]. The radius of gyration can be obtained from the slope by fitting the logarithm of the intensity $ln(I(q))$ versus q^2. However, the fit is only valid within the limit of $qR_g < 1.3$ [68].

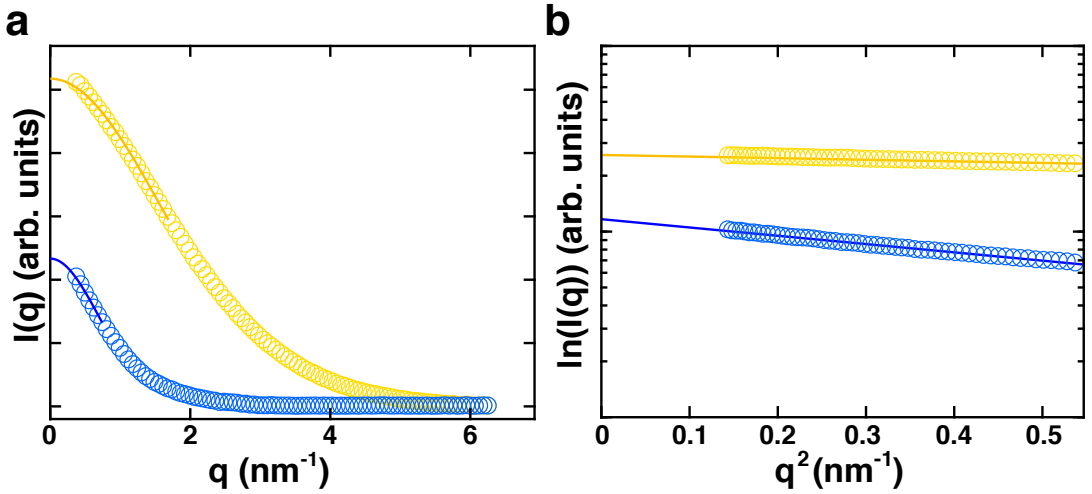

Figure 2.8: Guinier analysis and forward scattering. (**a**) Experimental scattering profiles for small (0.7 nm diameter) gold nanocrystals (yellow circles) and 15 bp DNA (blue circles). Corresponding guinier extrapolations to the $I(0)$ y-axis intercept for the gold nanocrystals (yellow line) and 15 bp DNA (blue line). (**b**) Corresponding scattering profiles in the Guinier representation with the scattering intensity $I(q)$ plotted versus the squared momentum transfer q^2. The number of s-bins shown for the experimental profile (colored circle) was reduced for clarity.

Moreover, the scattering intensity at zero momentum transfer $I(0)$ can be calculated via extrapolating the fitted line to the y-interception [31]. Since the forward scattering intensity $I(0)$ is directly proportional to the square of the scattering contrast, it is possible to determine the molecular weight (MW) of a macromolecule for a monodisperse solution by using the following relation [43]

$$I(0) = \kappa c(\Delta\rho)^2(MW)^2. \tag{2.59}$$

This equation includes the concentration of the macromolecules in solution c, the average electron density contrast $\Delta\rho$ and a proportional constant that can be determined from a measurement of a molecular weight standard (e. g. silver behenate, glassy carbon, or a protein of known molecular weight and shape) [43]. Furthermore, experimentally measured SAXS data are often adjusted by $I(0)$. Fig. 2.8 shows Guinier graphs for normalized experimental scattering profiles of 15 bp DNA and small gold nanocrystals (0.7 nm radius). While 15 bp DNA (~9 kDa) has a higher molecular weight than the gold nanocrystals (~6 kDa) [69], the gold provides a much higher electron density contrast which can be nicely seen at the higher y-axis intercept at $I(0)$.

2.4 *Ab initio* shape reconstruction

Another application using small-angle X-ray scattering scattering profiles, besides extracting structural parameters, is to reconstruct three-dimensional (3D) molecular shapes (see Fig. 2.9). This process is known as *ab initio* shape reconstruction since there is no pre-defined input structure [13]. During the past years, various reconstruction algorithms were published and great advances were achieved offering the possibility to model rigid mono-component systems. Those systems can be globular proteins or double stranded DNA (dsDNA), flexible molecules, or

protein-nucleic acids complexes [15, 16, 21, 23, 70]. Nevertheless, one important condition for a reliable *ab initio* shape reconstruction is the use of highly monodisperse samples and recorded data that do not suffer from experimental artifacts such as radiation damage, aggregation or boundary effects [31].

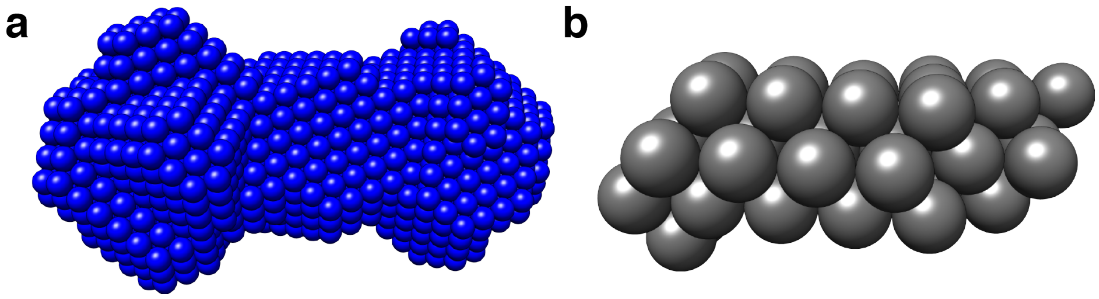

Figure 2.9: 3D bead models from *ab initio* shape reconstruction. (**a**) Reconstructed bead model of the signaling protein calmodulin. The configuration was calculated using *DAMMIF* [23]. (**b**) Reconstructed bead model of 20 bp DNA. The configuration was calculated using *SAXS3D* [16].

The *ab initio* determination of molecular shapes was first proposed by Stuhrmann [38, 39]. His approach was to consider a biomolecule as an object with uniform electron density and to calculate the scattering profile in terms of spherical harmonic expansions of the molecular surface. However, this approach was limited to molecules with globular shape and without cavities or multiple domains. In the late 1990s, Chacón *et al.* [14] and Svergun [15] published more advanced algorithms that offered the possibility to model more complex structures for the first time. This new approach approximated the outer molecular shape in form of the molecular electron density by a large number of dummy atoms [15], or dummy residues [21]. The number and arrangement of dummy scatterers is altered until the calculated theoretical scattering profile matches the experimental diffraction data. While some algorithms use a simulated annealing procedure [15, 21, 23] to approximate the experimental scattering profile, another possible approach is based on a "give'n'take" algorithm that was introduced by Walther *et al.* [16]. Typically, the bead model provides a low-resolution (1.5-2 nm) reconstruction of the molecule [18].

2.4.1 *DAMMIN* and *DAMMIF*

Two very popular programs for *ab initio* shape determination using finite elements (dummy scatterers) and a simulated annealing procedure are *DAMMIN* [15] and the improved version *DAMMIF* [23]. Both programs first evaluate the structural parameters of a molecule using the corresponding experimental SAXS profile. Then they start sampling from a search volume (see Fig. 2.10) with dimensions predefined by the maximum dimension D_{max} (*DAMMIN*) or the radius of gyration R_g (*DAMMIF*). The search volume is filled with densely packed beads (dummy atoms/residues) $N \gg 1$ of radius $r_0 \ll D_{max}$ to enclose the entire molecular shape [31]. The beads are then assigned either to be part of the particle phase (index=1) or the solvent phase (index=0) and are indexed accordingly. It is important to notice that the positions of the dummy beads follow a regular packing grid (see Fig. 2.10) and do not resemble the positions of atoms or residues within a real molecule. The shape of the molecule is given by the dummy atom model (DAM), which is described by a binary configuration vector X. The corresponding theoretical

scattering profile is calculated using spherical harmonics considering beads with index=1 [31]. This configuration vector X is then used to calculate the theoretical scattering profile of X using spherical harmonics (see Eq. 2.25), which provide good computational efficiency [71].

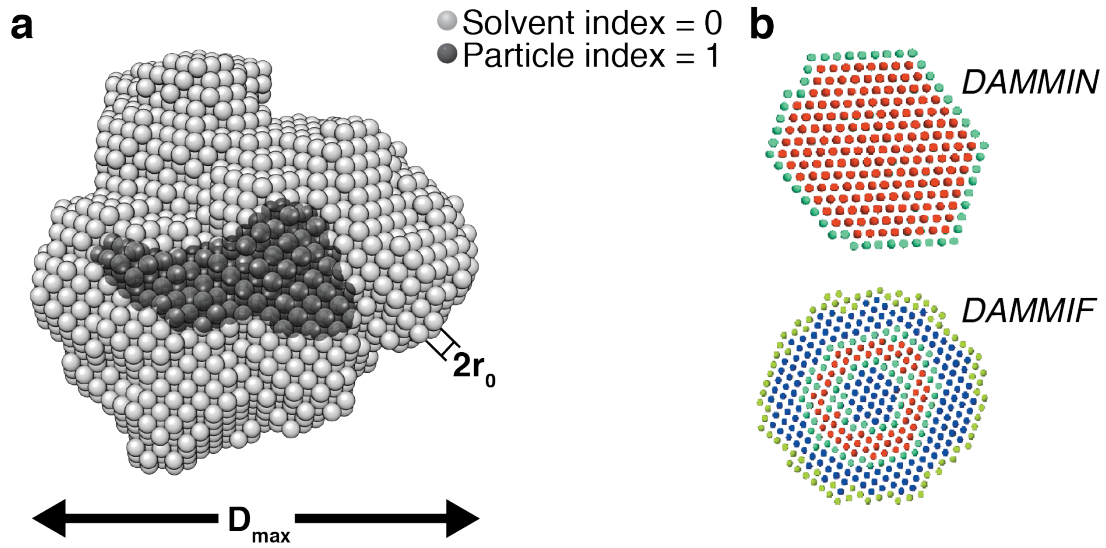

Figure 2.10: Initial search volume and dummy atom model of *DAMMIN* and *DAMMIF*. (**a**) Representation of the search volume and particle shape visualization in a bead model based on the *ab initio* algorithm defined by the maximum dimension of the molecule D_{max}. The light gray beads represent dummy atoms in the solvent state and dark gray spheres the particle state. Image inspired by Ref. [23]. (**b**) Cross-section of dummy atom models of *DAMMIN* (top) and *DAMMIF* (bottom). The different colors indicate either the particle (red) or the solvent phase (turquoise, blue and green) of the beads. While *DAMMIN* allows phase transition of the dummy atoms anywhere in the search volume, the change of state is restricted to red and turquoise beads in *DAMMIF*. In the bottom image, the *DAMMIF's* extensible mapped area is indicated by green solvent beads. Image reproduced from Ref. [23] with permission from the International Union of Crystallography.

The simulated annealing process aims to find a configuration X of dummy beads that matches the experimental SAXS profile best and minimizes the goal function $F(x)$ [15]

$$F(X) = \chi^2 + \sum_{i=1} \alpha_i P_i(X) \tag{2.60}$$

by using

$$\chi^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(\frac{I_{exp}(q_i) - cI_{calc}(q_i)}{\sigma(q_i)} \right)^2 . \tag{2.61}$$

χ^2 is the difference between the experimental $I_{exp}(q)$ and the theoretical calculated $I_{calc}(q)$ scattering profile with the experimental error $\sigma(q)$ at each q-value. Moreover, N the number of experimental points and c the scale factor. Furthermore, P_i describes penalty terms which are weighted by $\alpha_i > 0$. These enforce physical constraints (e.g. compactness and interconnectivity constraints) or can apply additional predefined structural information (e.g. symmetries) to guide the annealing process [23].

The minimization process is implemented as follows [15]:

1. The process starts from a initial random configuration X_0 at the so-called high annealing temperature $T = F(X_0)$.

2. A bead from configuration X_0 is randomly selected and its phase is changed to either the particle phase (index = 1) or the solvent phase (index = 0) to obtain the new configuration X' (see Fig. 2.10). The difference between the goal functions of both configurations is calculated: $\Delta = F(X') - F(X)$.

3. If the difference calculated in step 2 yields $\Delta < 0$, X' is instantly chosen as the new configuration; if $\Delta > 0$, X' is chosen with the probability $exp(\Delta/T)$. Step 2 is repeated with configuration X' if accepted or with configuration X if not.

4. The annealing temperature T is kept constant for $100 \times N$ reconfigurations (N is the number of beads in the search volume) or $10 \times N$ successful reconfigurations ($X \rightarrow X'$). If one of the criteria is fulfilled first, the annealing temperature is slightly cooled to $T' = 0.9 \cdot T$ reducing the probability to adopt a 'worse' fitting reconfiguration with $\Delta > 0$. Finally, the system is cooled until $F(X)$ cannot be minimized any further.

In contrast to a pure Monte Carlo based search algorithm, the simulated annealing process prevents the goal function $F(X)$ from getting trapped in local minima by also accepting reconfigurations that increase $F(X)$ with a certain probability (see step 3) and are 'worsening' the configuration [31].

While the search volume is fixed during the search procedure in *DAMMIN*, the volume is allowed to vary and extend in *DAMMIF* if necessary. Moreover, *DAMMIF* was designed to speed up the refinement during the simulated annealing process and reduce the shape determination from hours to minutes [23].

2.4.2 *SAXS3D*

Compared to the reconstruction approaches, *DAMMIN* and *DAMMIF*, which start from a predefined search volume, the program *SAXS3D* implements a "give'n'take" algorithm that places point scatterers on a hexagonally close-packed lattice with a specific constant lattice spacing. Similar to *DAMMIN* and *DAMMIF*, a theoretical diffraction pattern is calculated using spherical harmonic expansions (see Eq. 2.25) and a targeted scoring function F is minimized, to improve the fit between the theoretical calculated diffraction pattern of the model and the experimental SAXS profile [16].

The "give'n'take" procedure improves the fit between the experimental scattering profile and the theoretical calculated diffraction pattern with following operations[16]:

1. The process starts with a single bead $N = 1$ and a virtually infinite scoring function F.

2. The list of beads is first randomized and ordered according to ascending number of nearest neighbors. An additional bead is placed as immediate neighbor to the bead located on top of the list at a random, unoccupied position. The scoring function of the new arrangement F' with $N + 1$ beads is calculated. If $\Delta = F' - F < 0$, the new

configuration is accepted. Else, another possible neighboring lattice point is selected. If all positions fail to improve F, another bead is selected as new reference bead. If no bead is added, the configuration continues with the initial number of beads N.

3. Again, the list of $N + 1$ (or N) beads is randomized as described in the first step, the first bead from the list is removed and the score of the configuration is calculated. If the score has not improved ($\Delta > 0$), the bead is placed back and the algorithm proceeds with the next bead in the list. Otherwise if $\Delta < 0$, the removal is accepted.

4. The program iterates through the additions (step 2) and removals (step 3) until F cannot be minimized any further.

5. A randomly chosen bead is relocated from the original location to a new position as the nearest neighbor of another randomly selected bead. The relocation is performed on all beads until the first one is accepted because it satisfies the $\Delta < 0$ criterion. If a valid relocation was found, the algorithm proceeds with step 1. If no score-lowering move was found, the reconstruction terminates with the final model.

2.4.3 Normalized Spatial Discrepancy

However, the reconstruction of 3D shapes based on 1D scattering profiles do not provide unique solutions but a range of models with nearly identical fits to the experimental scattering profile [43]. Thus, several separate runs yield different geometrical solutions caused by the Monte Carlo based start [66]. Typically, multiple independent reconstruction runs are performed to obtain a range of models. Kozin and Svergun [72] introduced the normalized spatial discrepancy (NSD) as a quantitative comparison of similarity between three-dimensional objects.

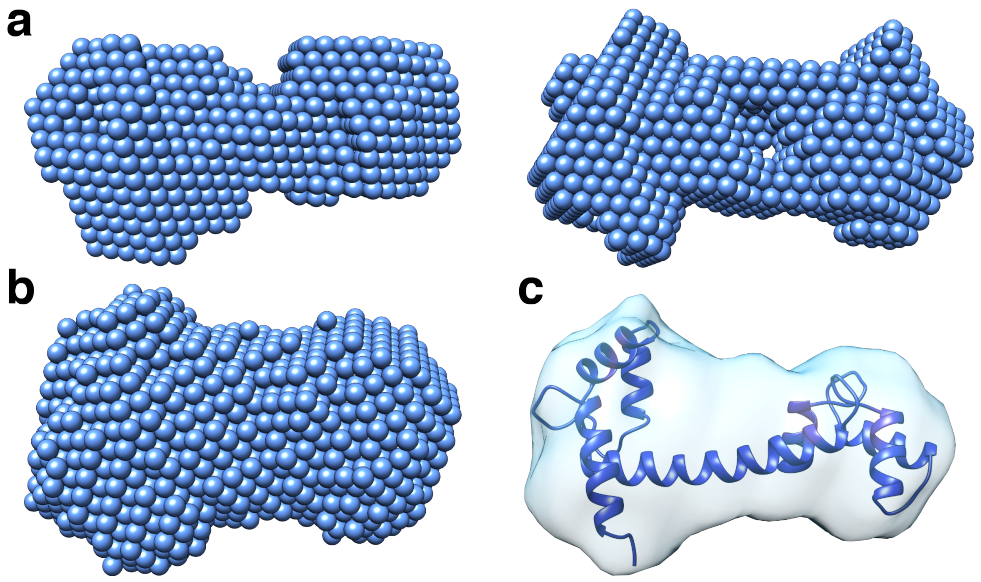

Figure 2.11: Example of *ab initio* shape determination of Calmodulin. (**a**) Two independent reconstruction models obtained from the same experimental data set using *DAMMIF*. (**b**) Superimposed and averaged model from 20 independent reconstruction runs created by *DAMAVER*. (**c**) Standard electron density map converted from the model in (b) with *SITUS*. PDB file (PDB code 1EXR [57]) obtained from Ref. [73].

Here, the beads in a first model are assigned to a point set S_1 and the beads in a second model are assigned to the point set S_2. For each point in S_1, the minimum value among distances between this point and all points in S_2 is determined. The same procedure is performed for all points in S_2 to the points in S_1. The resulting distances are added and normalized against the average distances between neighboring points from set S_1 and S_2 [31]. If the two models are almost identical, the NSD value tends to 0 whereas it exceeds 1 for objects which systematically differ from each other [72]. The obtained models can be weighted, superimposed and averaged (see Fig. 2.11) based on the NSD to yield the most probable model using an automated program such as *DAMAVER* [74]. This process can access the uniqueness of a structure and improve the robustness of the fit [43]. The final bead model can be converted into standard electron density map formats to dock high-resolution structures (see Fig. 2.11) inside the reconstructed model by a software called *SITUS* [75, 76].

2.5 Label Triangulation

It can be very challenging to obtain precise structural information of large biological macromolecules or big complexes formed from multiple molecular components. However, in 1972, Hoppe proposed that measuring the distances between a sufficient number of components within a complex can lead to a deduction of the three-dimensional arrangement and thus the quaternary structure using a novel technique denoted label triangulation [77, 78]. The idea is to chemically or genetically modify different subunits separately and the reconstruct the complex from its subunits. This structural technique is based on a technique for the determination of distances between heavy atoms by X-ray diffraction in solution.

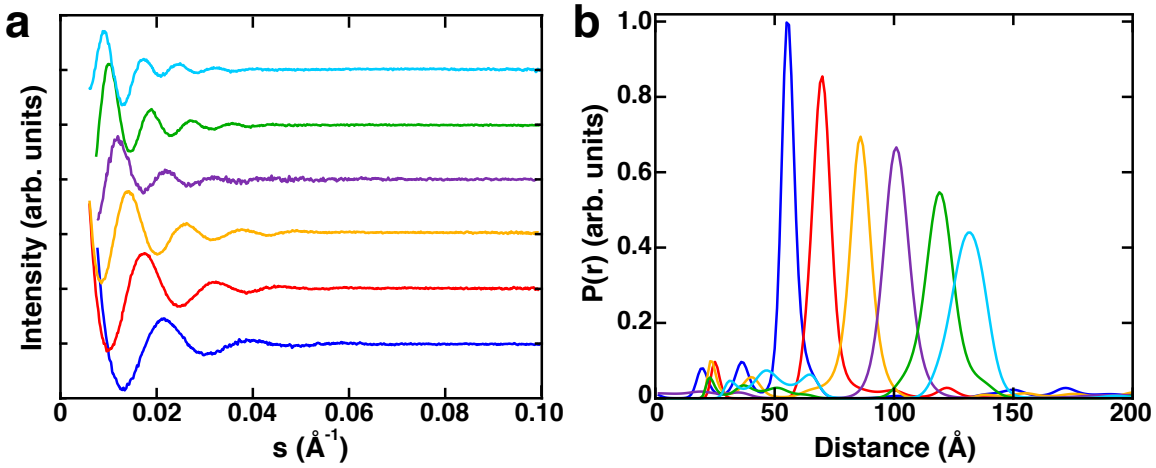

Figure 2.12: Distance distribution for double labeled DNA. (**a**) Scattering interference pattern obtained for 10 bp (blue), 15 bp (red), 20 bp (yellow), 25 bp (purple), 30 bp (green) and 35 bp (cyan) double labeled DNA duplexes. (**b**) Probability distance distribution $P(r)$ for the interference pattern from a). Distributions can be obtained via Fourier Transformation of the interference pattern. Reproduced from Ref. [69].

In fact, the first approach using heavy-atom markers dates back to 1947 by Kraty and Worthmann [79, 80]. They measured scattering profiles of small organic molecules using attached iodine labels. As described in Section 2.1.4, the scattering intensity is highly influenced by differences in electron density between solution and solvent. Thus the recorded signal is

dominated by the heavy-atom markers. Kraty and Worthmann were able to deduce the right inter-label distance after subtracting the scattering profiles of the single labeled sample from the traces obtained from double labeled molecules. Later, Vainshtein *et al.* [81] used mercabide markers to measure distances between heavy-atom markers in proteins. Moreover, the advance of modern synchrotron X-ray sources and the application of gold nanocluster markers (Fig. 2.12) reliably revealed entire distance distributions between pairs of labels [41, 82, 83].

The basic idea behind these methods is to extract the interference term $\Delta I(q)$ caused by the two labels (Fig. 2.12a). This interference term contains the desired distance information which can be obtained via Fourier transformation [77] (Fig. 2.12a,b):

$$\Delta I(q) = f_1 \cdot f_2 \cdot \frac{\sin(qr)}{qr}. \tag{2.62}$$

Here, f_1 and f_2 are the atomic scattering factors, the momentum transfer is q and r is the distance between the two electron rich particles with $r = |\mathbf{r_1} - \mathbf{r_2}|$. However, considering the flexible nature of biological molecules, the label-label distance cannot be assumed to be given by a single value but rather by a distribution $P(r)$ with a certain width (Fig. 2.12b). Thus, Eq. 2.62 can be rewritten to [69]:

$$\Delta I(q) = \sum_{r=0}^{r_{max}} P(r) \cdot f_1 \cdot f_2 \cdot \frac{\sin(qr)}{qr} \tag{2.63}$$

A more detailed description of the labeling and measuring procedures can be found in Chapter 4.

Chapter 3
Experimental Setup for SAXS

After having established basic theoretical concepts, which are relevant in small-angle scattering studies, this chapter briefly presents the main experimental requirements and characteristics of SAXS.

3.1 Introduction to Synchrotrons

After Wilhelm Conrad Röntgen discovered X-ray radiation at the Julius-Maximilians-University of Würzburg in 1895, the anode tube design remained the production source of X-ray waves for a long time [84]. An X-ray tube is designed in a way that electrons are emitted from a cathode and subsequently accelerated towards a cooled metal anode [85]. The anodes are usually made of copper (Co) or molybdenum (Mo) [65]. Electrons that collide with atoms in the target anode cause two types of radiation. First, a continuous spectrum of X-rays is emitted as bremsstrahlung since the electrons are decelerated [31]. Second, if the electron energy is sufficient to eject electrons within the inner shells (e.g. K shell) of the anode atoms (see Fig. 2.3), this vacancy is filled by an electron from a higher shell (L or M) [85]. As there is a distinct difference between the two energy levels, X-rays with well-defined line spectrums are emitted [85]. Although this fluorescent radiation is typically 50-100 times more intense than the bremsstrahlung, X-ray radiation is emitted in all directions leaving only a small fraction to target the sample [31].

However, in the 1970s, major improvements in form of bright synchrotrons were achieved (see Fig. 3.1). While these synchrotrons were exclusively available for high energy nuclear physics experiments in the beginning, the first facilities tailored to the needs of scattering experiments were established in the mid 1970s [31]. From that point on, several generations of synchrotrons helped to dramatically increase the beam intensity available, allowing for precise measurements of scattering patterns from biological macromolecules within the order of seconds [30, 86]. Additionally, they enabled experimentalists to tune the X-ray energy within a certain range or to focus the beam on a spot as small as several square nanometers [87, 88]. Nowadays, fourth-generation synchrotrons are available and the application of X-ray free-electron lasers (XFELs) provides highly collimated pulsed X-ray beams compared to third-generation sources [86]. Recently, the most powerful X-ray source ever built started to operate

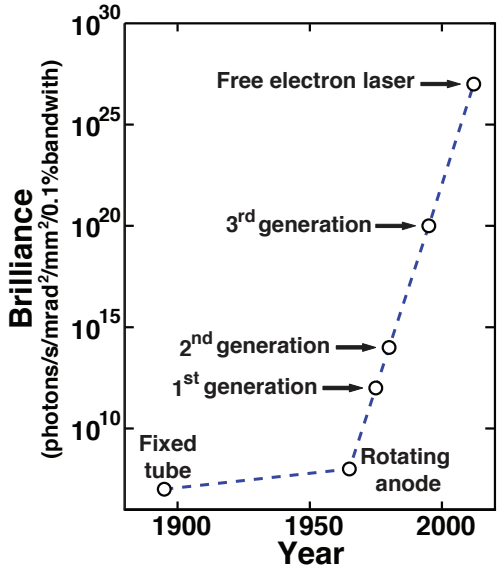

Figure 3.1: Historic development of the brilliance of X-ray sources. The term synchrotron brilliance is defined in Section 3.2. The peak brilliance of free electron lasers exceeds the one of common third-generation sources by a large factor. Figure adapted from Ref. [30] with permission from Wiley. Copyright ©2011 John Wiley & Sons, Ltd..

in Hamburg [89]. The European X-ray Free Electron Laser (European XFEL) provides electron energies up to 17.5 GeV and a peak brilliance of 5×10^{33} photons/s/mrad^2/mm^2/0.1% bandwidth which is higher by a factor of $\sim 10^{25}$ compared to earlier fixed anode X-ray tubes (Fig. 3.1) [90, 91].

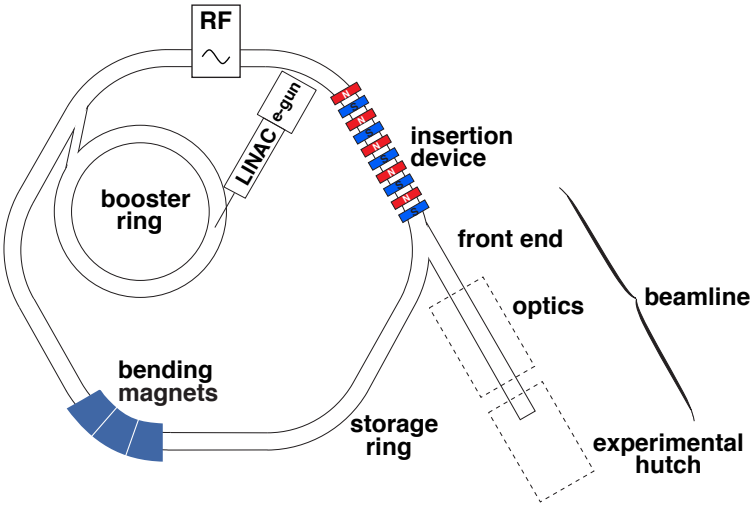

Figure 3.2: Schematic of the key components of a modern synchrotron facility. The figure is adapted from Ref. [91] with permission from Wiley. Copyright ©2011 John Wiley & Sons, Ltd..

The source of X-rays in synchrotron facilities are charged particles (e.g. electrons) moving at relativistic speed on an almost circular trajectory within an evacuated storage ring [92]. In today's synchrotron facilities, a linear accelerator (LINAC) and a small booster ring helps to pre-accelerate bundles of particles before injecting them into a large storage ring (Fig. 3.2) [85].

One important parameter for synchrotrons is the energy of the charged particles within that storage ring, as it sets the energy range a source can provide to the experimentalist [91]. For example, the European Synchrotron Radiation Facility (ESRF) operates with an electron current of 100 mA at an energy of 6 GeV and a storage ring of 844 m circumference [93]. The storage ring is one of the key components and usually consists of curved parts with bending magnets to guide the charged particles on a closed orbit and additionally to straight sections (Fig. 3.2) [91]. Within these straight segments, insertion devices (wigglers or undulators) are placed [30]. They force the charged particles to undergo small-amplitude oscillations. As a result of accelerated charged particles in the curved and straight sections, the so-called synchrotron radiation is emitted [31]. After these X-rays are filtered by a monochromator to select for a narrow wavelength bandwidth, they are focused and aligned via special mirrors and lenses [94]. Finally, the beam is delivered to the experimental hutch and directed to expose the desired sample.

3.2 Properties of the X-ray Beam

There are several properties describing the quality of a synchrotron beam [95]. First, there is the number of photons the source can produce per second. Second, the beam diverges on its trajectory, although the X-rays are emitted much more collimated in bending magnets or insertion devices compared to X-ray anodes. The convention is to quote the spread of the beam in mili-radian (mrad) in both vertical and horizontal direction [85]. Third, as the beam is directed on the targeted sample, an important parameter for the quality of the facility is the source area giving the smallest spot onto which the X-ray beam can be focused (commonly in square millimeters mm^2) [95]. The last property is the unit bandwidth allowing to include a factor for the range of photon energies contributing to the intensity. Usually, the unit bandwidth (BW) is chosen as 0.1% of the relative energy bandwidth [85]. Together, all these parameters combined allow the comparison between X-ray sources. This joined value is denoted brilliance and is given as [91]:

$$\text{Brilliance} = \frac{\text{photons/second}}{(\text{mrad})^2(\text{mm}^2\text{source area})(0.1\%\ \text{bandwidth})} \tag{3.1}$$

3.3 X-ray Sources

There are multiple sources of synchrotron radiation at a synchrotron facility. The following section outlines three devices used to create X-ray beams.

3.3.1 Bending Magnets

As previously described, X-ray waves propagate in all directions from anode sources. In contrast, synchrotron radiation is emitted in a narrow cone by charged particles traveling close to the speed of light (Fig. 3.3a) [85]. One radiation source are bending magnets (BM), which are primarily used to keep the charged particles on a closed orbit. However, they also produce so-called bending magnet radiation which covers a continuous spectrum of X-ray energies (Fig. 3.3b) [91]. The radiation is emitted in tangential direction and forced into a narrow cone

by the Lorentz transformation and the Doppler effect (Fig. 3.3a) [31]. The angle ψ, denoting the degree of the beam's divergence is given by [92]

$$\psi \approx \frac{m_e c^2}{E} \approx \gamma^{-1}. \tag{3.2}$$

In this equation, the rest mass of electrons is m_e, the velocity of light in vacuum is c, the relativistic energy is E (storage ring energy) and the Lorentz factor is γ. The radiation is linearly polarized to an observer in the orbit plane of the bending magnets (Fig. 3.3a). If viewed from above, it looks elliptically polarized [85]. The other parameters are the characteristic frequency ω_c and the corresponding critical energy E_c dividing the emitted power spectrum in a way that one half is radiated below and the other half is radiated above E_c (Fig. 3.3). The critical energy is given by [91]

$$E_c[\text{keV}] = \omega_c \hbar = 0.665 \cdot E^2[\text{GeV}] \cdot B[\text{T}]. \tag{3.3}$$

Here, E is the energy of the charged particles in the storage ring and B the magnetic field of the bending magnet [92]. In addition, the radiated power from a bending magnet of length L and a magnetic field B located in a storage ring with an electron current I at energy E is expressed as [91]

$$P[\text{kW}] = 1.266 \cdot E^2[\text{GeV}] \cdot B^2[\text{T}] \cdot L[\text{m}] \cdot I[\text{A}]. \tag{3.4}$$

One example for a beamline receiving the X-ray beam from a synchrotron bending magnet is BM29 at ESRF [96].

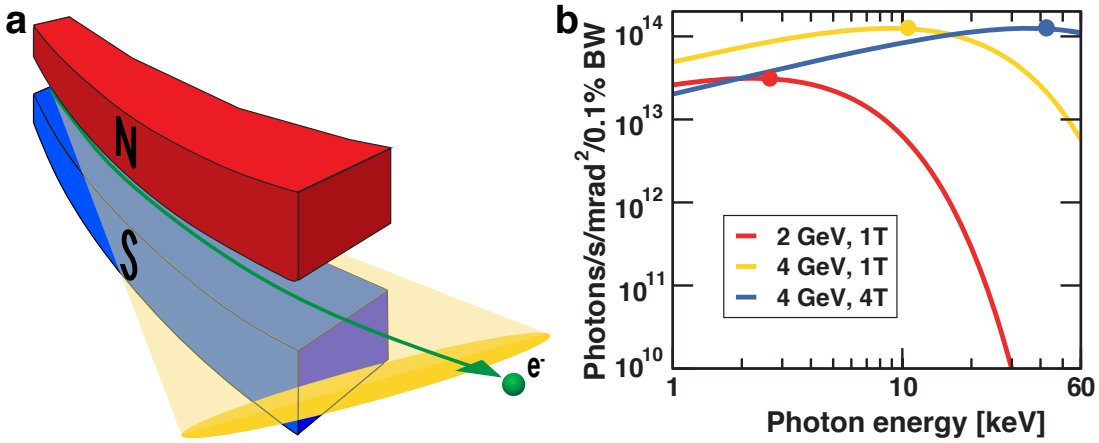

Figure 3.3: Synchrotron bending magnets. (**a**) Schematic of a bending magnet and the emitted linearly polarized radiation cone by an electron. (**b**) Continuous X-ray spectrum produced by a bending magnet as a function of photon energy. The critical energy E_c depicted as circles is given by Eq. 3.3. Figures are reused from Ref. [91] with permission from Wiley. Copyright ©2011 John Wiley & Sons, Ltd..

3.3.2 Inserting Devices

Another source of synchrotron X-rays are so-called insertion devices (ID) placed within the straight sections of a storage ring. Insertion devices are built-up by a linear series of dipole magnets forcing the charged particles to execute an oscillatory trajectory [91] (see Fig. 3.4a,b).

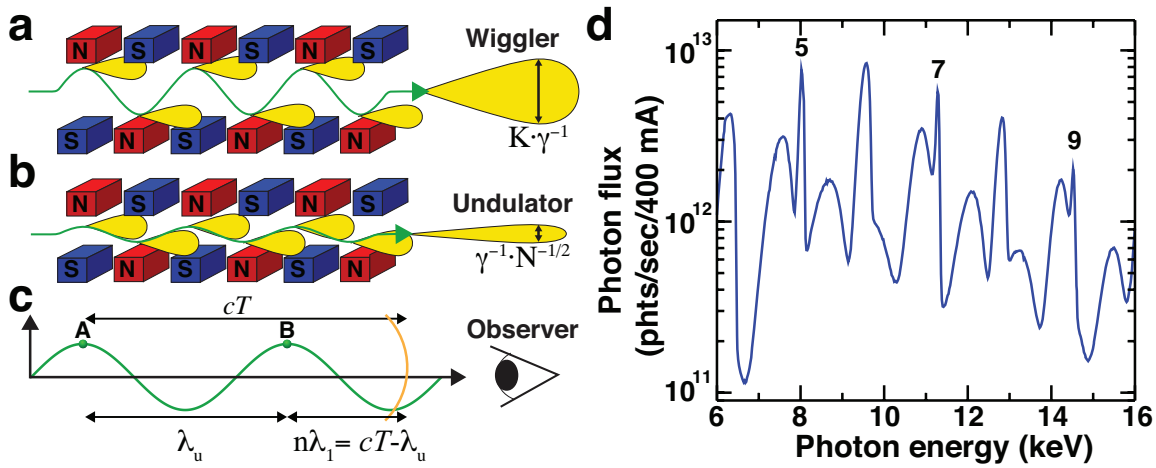

Figure 3.4: Synchrotron insertion devices. Schematic of wigglers (**a**) and undulators (**b**) differing in the amplitude of the oscillating trajectory forced by a series of magnets. The amplitude of the oscillations is greatly exaggerated in this figure. The figure was inspired by Fig. 3.16 in Ref. [91]. (**c**) Constructive interference occurs when the separation between waves emitted from equivalent points A and B on the oscillating path corresponds to multiples of the fundamental wavelength λ_1. The figure was inspired by Fig. 2.8 from Ref. [85]. (**d**) Photon flux provided by a in-vacuum undulator at the Swiss Light Source (SLS). Sharp 5^{th}, 7^{th} and 9^{th} harmonic energy lines are labeled. Figure is adapted from Ref. [91] with permission from Wiley. Copyright ©2011 John Wiley & Sons, Ltd..

There are two types of insertion devices: wigglers and undulators [92]. In contrast to bending magnets, both wigglers and undulators are purely designed for the emission of high brilliance beams [95]. The major difference between both insertion devices is that the magnetic field is higher in a wiggler device [85]. Thus, the amplitude of the oscillations is larger compared to undulators [31] (Fig. 3.4a,b). On the one hand, since the amplitudes of the oscillations are larger than the opening angle γ^{-1} of the X-ray beam, the emitted radiation from each dipole magnet pair ('wiggle') adds incoherently in wigglers producing a pulsed intensity beam. On the other, hand the small-amplitude oscillations on the order of γ^{-1} in undulators allow for constructive interference of waves from each oscillation to a longer pulse [85]. Comparing both insertion devices, wigglers provide an increase in brightness by a factor of N where N is the number of periods (number of magnet poles) and a cone with an opening angle $\sim K\gamma^{-1}$. In contrast, undulators can increase the brightness by a factor of N^2 and a much smaller opening angle $\sim\gamma^{-1} \cdot N^{-1/2}$ [31] (Fig. 3.4a,b). Similar to the power of a bending magnet, the total radiated power by a wiggler is given by [92]

$$P[\mathrm{kW}] = 0.633 \cdot E^2[\mathrm{GeV}] \cdot B^2[\mathrm{T}] \cdot L[\mathrm{m}] \cdot I[\mathrm{A}]. \tag{3.5}$$

In this equation, B and L are the maximum magnetic field and length of the wiggler. I and E are the electronic current and energy of the storage ring.

While the energy spectrum of a wiggler is comparable to the one of a bending magnet, an undulator devise emits a spectrum with distinct maximums [92]. Considering the phase difference between photons in an undulator, narrow maximums occur in an equally-spaced pattern as they are caused by constructive interference (Fig. 3.4c,d) between each undulation [85]. The

energy of photons produced from an undulator is therefore [91]

$$E_n[\text{keV}] = 0.95 \cdot \frac{n \cdot E^2[\text{GeV}]}{(1 + K^2/2)\lambda_u[\text{cm}]} \tag{3.6}$$

with the number of harmonics n, the total particle energy E, the period of oscillation within the device λ_u and the unit-less K-factor

$$K = \frac{eB}{2\pi mc}\lambda_u. \tag{3.7}$$

Here, m is the rest mass of the particle, c is the velocity of light and B is the maximum magnetic field. Thus, the photon energy emitted from an undulator can be tuned by varying K or B respectively. Based on their composition wigglers have $K \gg 1$ (~20) while undulators and their weak magnetic fields result in $K \ll 1$ [31, 85]. An example for a beamline operating with a wiggler X-ray source is beamline 4-2 at the Stanford Synchrotron Radiation Light-source (SSRL) [97]. Beamline 12-ID at the Advanced Photon Source (APS) in contrast, works with an undulator [98].

3.4 SAXS Beamline

For experimentalists, the beamline is the most important component of a synchrotron facility since every beamline is tailored to a certain set of experiments. Fig. 3.5 illustrates the arrangement of the bio-SAXS beamline BM29 at the European Synchrotron Radiation Facility [96].

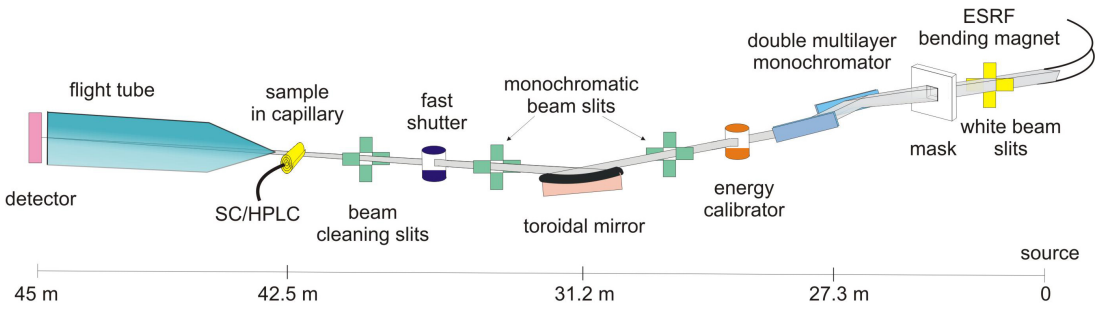

Figure 3.5: Schematic of the bio-SAXS beamline BM29 at the European Synchrotron Radiation Facility displaying the X-ray optics to filter, focus and align the X-ray beam. Furthermore, the sample stage and detector are shown. Adapted from Ref. [96]

The X-rays produced from a bending magnet or an insertion device are first masked, filtered and directed by optical devices (Fig. 3.5). Second, a narrow band of wavelengths is selected by a monochromator and finally the beam is cleaned, focused and aligned onto the sample. The last part of a beamline is the experimental hutch where the sample is loaded into a quartz glass capillary and the scattering data are recorded at a certain detector distance after passing an evacuated flight tube (Fig. 3.5). By tuning the detector distance, one can vary the recorded q-range of the diffraction signal. In case of the BM29, the recordable q-range is 0.025-5 nm^{-1} with a detector distance of 2.867 m at a photon energy of 12.5 keV [96]. A more detailed description of the measurement protocol can be found in Chapter 4 and the Materials and Methods section of Chapter 5.

3.4.1 Monochromator

Commonly, experiments performed at a synchrotron require a specific wavelength or, in the case of anomalous studies such as X-ray Absorption Spectroscopy (XAS) [94], Multi-wavelength Anomalous Diffraction (MAD) [49] or anomalous small-angle X-ray scattering (ASAXS) [2, 43, 55], tuning of the energy. Moreover, since scattering data are recorded as a function of momentum transfer q which varies with the wavelength of the incident photons ($q = 4\pi \sin(\theta/\lambda)$), the obtained signal gets blurred if the energy spectrum is too broad. Since bending magnets and wigglers produce a broad spectrum of wavelengths and even radiation from undulators is not sufficiently monochromatic, single crystal monochromators are employed to the X-ray beam [99]. These devices narrow the band of wavelengths of the beam by diffraction based on Bragg's law [100]

$$m\lambda = 2d \cdot \sin(\theta) \tag{3.8}$$

to filter out a sharp band of wavelengths. Here, m is the order ($m = 1, 2, 3, \ldots$), λ the wavelength, θ the incident angle and d the lattice spacing (Fig. 3.6a). Usually, an energy resolution of $\Delta E/E \leq 10^{-4}$ can be achieved [99]. The standard method to monochromatize a X-ray beam is to use two perfect crystals separated by the distance D (Fig. 3.6b). While rotating the device, the vertical offset Δy of the beam is changed until the desired photon wavelength is reflected at a certain height h (Fig. 3.6b) [94].

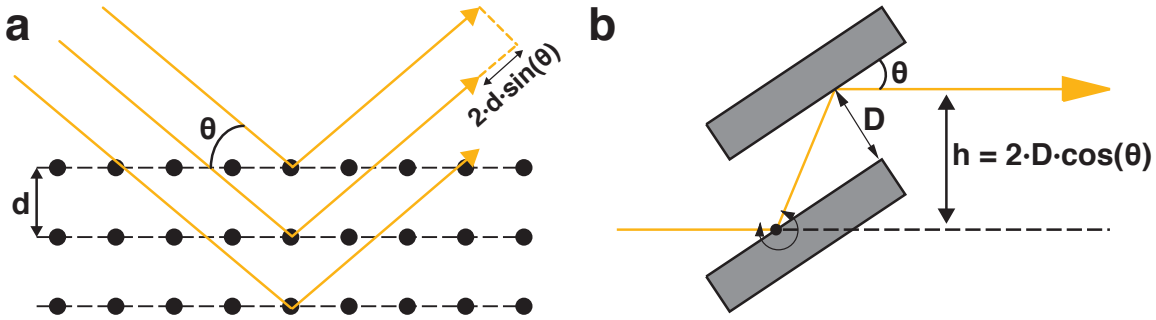

Figure 3.6: Principle of a double crystal monochromator based on Bragg's law. (**a**) Bragg diffraction on a crystal with lattice spacing d. The beams with incident angle θ have a phase difference of $2 \cdot d \cdot \sin(\theta)$ between the different crystal planes. (**b**) Geometry of a double-crystal monochromator with distance D between the two crystals. The graphical representation was inspired by Fig. 3.5 of Ref. [99].

Part II

Results

Chapter 4

Recording and Analyzing Nucleic Acid Distance Distributions with X-ray Scattering Interferometry (XSI)

Summary

Most structural techniques provide averaged information or information about a single predominant conformational state. However, biological macromolecules typically function through series of conformations. Therefore, a complete understanding of macromolecular structures requires knowledge of the ensembles that represent probabilities on a conformational free energy landscape. Here we describe an emerging approach, X-ray scattering interferometry (XSI), a method that provides instantaneous distance distributions for molecules in solution. XSI uses gold nanocrystal labels site-specifically attached to a macromolecule and measures the scattering interference from pairs of heavy metal labels. The recorded signal can directly be transformed into a distance distribution between the two probes. We describe the underlying concepts, present a detailed protocol for preparing samples and recording XSI data, and provide a custom-written graphical user interface to facilitate XSI data analysis.
NOTE: Use ultrapure water in all solutions and protocol steps.

4.1 Introduction

Richard Feynman famously said "… everything that is living can be understood in terms of the jiggling and wiggling of atoms" [101]. Biological macromolecules such as unfolded or partially folded RNAs or intrinsically disordered proteins are especially dynamic, given the noncovalent forces that hold them together, their aqueous surroundings, and physiological temperature that provides thermal energy. Moreover, significant conformational changes of

This Chapter submitted by Zettl *et al.* [1] to *Current Protocols in Nucleic Acid Chemistry* and is currently in press. I designed the study, performed the experiments, analyzed the data and contributed to writing the manuscript with help of the other authors. The chapter is adapted with permission from the John Wiley & Sons.

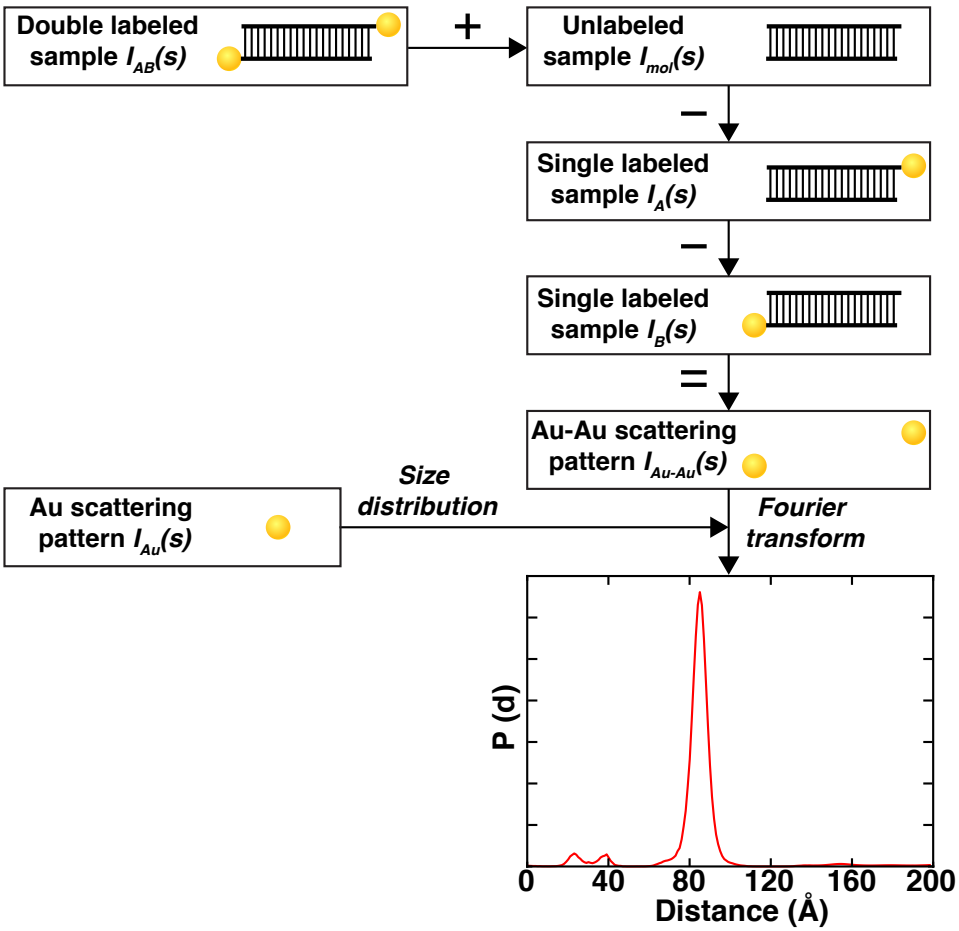

Figure 4.1: Schematic of AuSAXS workflow to determine gold label-gold label distance distributions. Scattering intensity equation for a single-labeled molecule. The scattering signal can be decomposed into a sum of the individual scattering contributions: double-labeled sample, the macromolecule only, two gold-macromolecule cross-terms, and the interference from the gold labels. Schematic of the workflow to determine the Au-Au distance distribution. The SAXS profiles of the shown samples are used to extract the gold-gold interference scattering profile. The interference pattern is Fourier transformed into a distance distribution using basic profiles generated with the size distribution of Au nanocrystals.

molecules can be triggered by external stimuli and are typically integrally involved in the functions of biomolecules [102–104]. Thus, conformational changes play an important role in understanding the basic mechanics and are key in reconstructing causes from the molecular level to macromolecular systems.

This chapter describes an emerging molecular ruler, termed X-ray scattering interferometry (XSI) [69, 82, 83, 105–108], which can be used to generate whole distance distributions at Ångström resolution. XSI measures the interference of scattered X-rays between two specifically attached gold nanocrystals (Fig. 4.1). The strengths of XSI are that it provides: *i*) distance information in solution; *ii*) the distance information that is unperturbed by temporal averaging because scattering is fast relative to atomic motions; *iii*) the direct mathematical relationship between scattering and distance, formally related by Fourier transformation, allows XSI data to be unambiguously converted into a calibrated distance distribution; and *iv*) while sample preparation is time-consuming, it is straightforward, highly reliable and highly reproducible.

Over the past decade, XSI has been successfully applied to nucleic acids and nucleic acid/protein complexes [41, 69, 82, 83, 105–108]. Labelling strategies employing gold nanocrystals of various sizes, diverse macromolecules, and variable attachment positions have been reported: So far, demonstrated labeling strategies include: *i*) end labeled DNA nanocrystal conjugates [69, 82, 83, 105, 106, 109, 110], *ii*) end labeled RNA molecules [83, 107, 108], *iii*) gold labels positioned at defined internal sites of DNA or RNA helices [82, 83, 105–108], and *iv*) protocols to form single-labeled protein constructs [111–113].

This chapter covers the design and preparation of end-labeled nucleic acid samples. Below we present the procedures required to prepare samples, to acquire XSI data, and to generate ensemble distance distributions.

Section 4.2 describes sample preparation for end-labeled nucleic acid gold conjugates and includes a protocol for gold nanocrystal synthesis.

Section 4.3 describes the acquisition of a full data set at a synchrotron radiation facility for XSI analysis.

Section 4.4 describes the data analysis of XSI data and the use of a custom-written graphical user interface (GUI) in MATLAB.

The detailed protocols and the user interface presented in this chapter will enable scientists interested in molecular distance measurements to perform and analyze XSI measurements easily.

4.2 Basic protocol 1, Sample preparation

In this chapter, we focus on the preparation of end labeled nucleic acid samples for XSI. An additional protocol on labeling proteins will be forthcoming.

Briefly, thioglucose protected gold nanocrystals are synthesized using the method of Schaaff and coworkers [114]. DNA or RNA oligonucleotides are ordered with a commercially available C3-thiol-modification for end labeling, or a C2 dT amino-modification for internal labeling. High-performance liquid chromatography (HPLC), performed either in-house or by the oligonucleotide vendor, is used to purify the oligonucleotides. In the case of the internal C2 dT amino-modification, the amino group is converted to a thiol group using the commercially available SPDP (N-succinimidyl3-[2-pyridyldithio]-propionate) cross-linker. For a detailed protocol of internal label attachment see [83]. The gold nanocrystals couple to thiol groups, forming stable conjugates. A second HPLC purification step is used to purify 1:1 nanocrystal nucleic-acid conjugates, eliminating nanocrystals coupled to multiple oligonucleotides and excess gold particles. Finally, modified and unmodified single-stranded molecules are mixed in various combinations to form a sample quartet, which consists of one unmodified construct, two complementary single-labeled molecules with a single gold nanocrystal attached to one of the two labeling sites respectively, and one double-labeled construct. After HPLC purification and desalting, these duplexed constructs can be stored at −20 °C for several months. A full set of samples is required for the data analysis to work as explained in detail in Section 4.4.

CAUTION: Some of the chemicals and reagents used are flammable. Refer to material safety data sheets prior to use. Some of the reactions should be conducted in a well-ventilated fume hood. Use of personal protection equipment is highly recommended.

Materials

- Dithiothreitol (Thermo Fisher, cat. No. R0861)
- Tris hydrochloride
- FPLC cleaning solution (100 mM DTT, 20 mM Tris-HCl, pH 8.0)
- Ammonium acetate (Sigma Aldrich, cat. No. A1542)
- Size exclusion running buffer (150 mM ammonium acetate, pH 5.6)
- Methanol
- Acetic acid
- Hydrogen tetrachloroaurate (III) hydrate (Sigma Aldrich, cat. No. 50790)
- 1-thio-β-D-glucose (Sigma Aldrich, cat. No. T6375)
- Sodium borohydride
- Wet ice
- 3'-thiol modified oligonucleotides
- Oligonucleotides
- 2 M ammonium acetate, pH 5.6
- Sodium chloride (NaCl)
- Sodium hydroxid (NaOH)
- Low salt borate buffer (10 mM NaCl, 20 mM Na-borate, pH 7.8)
- High salt borate buffer (1.5 M NaCl, 20 mM Na-borate, pH 7.8)
- Ethanol
- Magnesium chloride ($MgCl_2$)
- Low salt acetate buffer (10 mM NaCl, 20 mM ammonium acetate, pH 5.6)
- High salt acetate buffer (1.5 M NaCl, 20 mM ammonium acetate, pH 5.6)

- G25 column, 26/10 Housing (Sigma Aldrich, cat. No. GE17-5087)
- FPLC system (Dual Wavelength detector recommended)
- Superdex 30 column, 16/600 Housing (Sigma Aldrich, cat. No. GE28-9893-31)
- 250 mL round-bottom flask
- Parafilm
- Magnetic stirrer/hotplate
- Vortexer
- 0.22 µm syringe filter units
- 3 kDa and 10 kDa Amicon spin filtration units (Sigma Aldrich)
- Dionex DNAPac Pa200 column, 9/250 Housing (Thermo Fisher, cat. No. 063421)
- HPLC System (Dual Wavelength detector recommended)
- Rotary evaporator equipped with a dry ice condenser and connected to an oil pump
- −20 or −80 °C freezer
- Nanodrop ND-1000 spectrophotometer or other UV spectrophotometer

The protocol below describes: *i*) the synthesis of monodisperse, thiol-passivated gold nanocrystals with 0.7 nm radius; *ii*) the preparation of end-labeled gold-oligonucleotides conjugates; and *iii*) the preparation of a sample quartet that is ready for XSI data acquisition. The concentrations cited below are based on ordering a 200 nmol scale nucleic acid synthesis, and can be adjusted for alternate quantities of starting material.

Synthesis and Purification of gold nanocrystals

Steps 1-2: Preparation of FPLC columns (start two days in advance)

1. Using a flow rate of 2 mL/min, wash the G25 column with two column volumes of water, followed by four column volumes of the FPLC cleaning solution (20 mM Tris-HCl, pH 8.0, and 100 mM DTT). Equilibrate the column with four column volumes of water.

2. Using a flow rate of 0.75 mL/min, wash the Superdex 30 column with two column volumes of water, followed by 5 column volumes of the FPLC cleaning solution (same cleaning solution as in step 1). Remove the FPLC cleaning solution with 5 volumes of water. Finally, equilibrate the Superdex 30 with two column volumes of the size exclusion running buffer (150 mM ammonium acetate, pH 5.6).

   Note: The gold nanocrystals cause the column resin to turn brown. This coloration is reversed by the DTT in the cleaning solution. If the resin does not revert to off white after 5 column volumes, apply additional cleaning solution before equilibrating the column. Start the cleaning and equilibration of both columns at least one day before the synthesis.

Steps 3-11: Synthesis of gold nanocrystals (4 hours)

3. Rinse a 250 mL round-bottom flask with isopropanol, dry in a heated oven and cool (to room temperature). Add a stir bar, cap the flask with parafilm and mount the flask above a magnetic stir plate.

4. Prepare 72 mL of a 5:1 v/v mixture of methanol-acetic acid (60 mL : 12 mL).

5. Weigh 0.544 g of hydrogen tetrachloroaurate (III) hydrate and immediately transfer it to the round bottom flask. Add 36 mL of the 5:1 methanol-acetic acid solution to the flask. The color should be a clear, bright orange.

6. Dissolve 1 g of 1-thio-β-D-glucose in 36 mL of the 5:1 methanol-acetic acid mixture and vortex the solution until the powder is fully dissolved.

7. Add the dissolved 1-thio-β-D-glucose to the 250 mL round-bottom flask. The solution should turn cloudy. Stir the mixture for 20 min at room temperature.

8. Weigh 0.9 g of sodium borohydride and dissolve it in 20 mL of ddH_2O by vortexing.

9. Carefully add the sodium borohydride solution dropwise to the reaction flask over 12-15 min. Use a 1 mL pipette for this step, or alternatively an addition funnel with a metering valve (for example Chemglass, cat. No. CG-1714).

   Important!: it is critical to keep the addition speed and droplet volume constant to ensure monodisperse and high-quality gold nanocrystals.

10. Stir the mixture for 30 min at room temperature.

11. Concentrate the reacted solution (approximately 92 mL) to 12-20 mL using a rotary evaporator. The water bath needs to stay at room temperature.

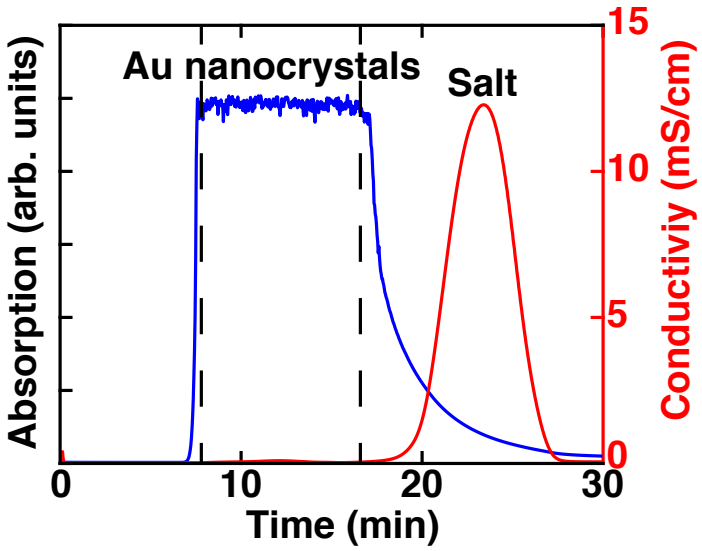

Figure 4.2: A sample FPLC chromatogram of Au nanocrystal desalting. The absorption of the Au nanocrystals (blue) is monitored at 260 nm and separated from the salt front (red) observed by a peak in conductivity. The sample eluting between the two dashed lines was used.

Steps 12-16: Purification of gold nanocrystals (10 hours)

12. Filter the crude nanocrystal solution using a 0.22 µm filter unit, and store the solution on ice.

13. Desalt the filtered nanocrystals with the prepared G25 column. Split the total volume into 2-3 runs. Load one aliquot of the solution (at most 5 mL with a standard sample loop) while applying a flow rate of 2 mL/min using water as the running buffer. Set the FPLC system up to monitor 260 nm or 280 nm to detect the Au nanocrystals. The particles should elute in a window between 8-14 min (see Fig. 4.2). Also monitor the conductance of the solution to detect the salt peak, which elutes after the nanocrystals. Repeat the desalting procedure for the additional aliquots. The desalted Au nanocrystals are stable can be stored at 4 °C for up to a few days or at −20 °C for months.

14. Concentrate the desalted Au nanocrystal solution to 10 mL with centrifugal filter units (3 kDa cutoff, 15 mL, 3,000×g, 40 min) at 4 °C. The flow through should be clear.

15. Purify a monodisperse population of Au nanocrystals on the Superdex 30 size exclusion column (see Fig. 4.3). Set up the FPLC system to monitor 260 nm or 280 nm to detect the nanocrystals. You can inject up to 5 mL of sample per run. After loading an aliquot of the filtered, desalted and concentrated nanocrystal solution, apply the size exclusion running buffer at 0.75 mL/min for at least 210 minutes. Collect only the center of the largest Au nanocrystal elution peak (see Fig. 4.3), and discard the lower and upper shoulders. Immediately desalt the solution using ddH_2O and centrifugal filter units (3 kDa cutoff, 15 mL) at 3,000×g and 4 °C. Repeat the centrifugal desalting three times and pool the concentrated particles after the final run.

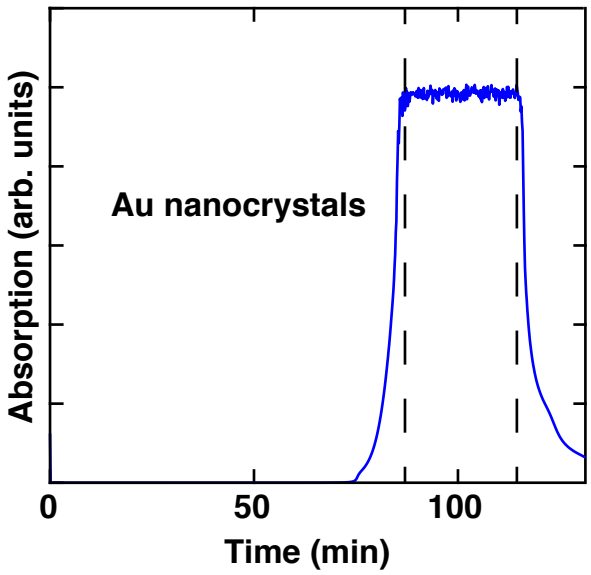

Figure 4.3: A sample FPLC chromatogram of Au nanocrystal size exclusion. The absorption of the Au nanocrystals (blue) is monitored at 260 nm. The sample between the two dashed lines can be used for a highly uniform particle size distribution.

16. Determine the final gold nanocrystal concentration by measuring the UV absorption. The extinction coefficient is 0.076 µm/cm at 360 nm. Store the solution at −20 °C. Typically one synthesis yields 3-10 µmol of purified gold nanocrystals.

Preparation of gold nanocrystal-nucleic acid conjugates

Step 17-22: Preparation of oligonucleotides (about 8 hours for 4 oligonucleotides)

17. Purchase 3'-thiol modified oligonucleotides (C3-S-S) and unmodified oligonucleotides with the same sequences. It is critical to use the short three-carbon linker for end labeled samples to ensure high-quality results for the measurements of the distance distributions. Design the construct such that the terminating bases are GC base pairs to minimize end fraying.

18. Purify the ordered oligonucleotides using the Dionex DNAPac 200 and anion exchange HPLC. Inject up to 100 nmol of oligonucleotide onto the column and apply a flow of 3 mL/min. The salt gradient is formed from low salt borate buffer (10 mM NaCl, 20 mM Na-borate, pH 7.8) and high salt borate buffer (1.5 M NaCl, 20 mM Na-borate, pH 7.8). Tune the period of the gradient according to the length of your oligonucleotide. Perform an analytical run before the preparative ones to determine the elution time of the product. For analytical runs adjust the injected sample according to the instrument detection sensitivity.

19. Desalt the purified oligonucleotides by buffer exchange into water using centrifugal filter units and ddH_2O (3 kDa cutoff, 4 mL, 4,000×g in a swinging basket centrifuge or 7,500×g in a fixed angle rotor, 30 min, 4 °C). Repeat this step three times. Finally, reduce the volume to ~40 µL using centrifugal filter units (3 kDa cutoff, 0.5 mL, 14,000×g using a

benchtop centrifuge, 30 min, 4 °C). Typically, ~60 nmol of oligonucleotide remain after purification of a 200 nmol scale synthesis. To protect against loss of oligonucleotide from a broken filter unit, keep the flow through and check its absorbance at 260 nm. The oligonucleotides can be stored in a freezer at −20 °C.

Steps 20-26: These steps use the thiol-modified oligonucleotides (about 3 hours)

20. Immediately before coupling oligonucleotides to gold nanocrystals, ensure that the pendant thiols are fully reduced by adding 150 µL of 200 mM DTT, 50 mM Tris-HCl, pH 9.0 and incubating for 30 min at 60 °C for DNA or 50 °C for RNA.

    Note: The DTT can be replaced by immobilized TCEP Reducing Gel (Thermo Fisher, cat. No. 77712). Follow the manual provided by the vendor to reduce and extract the oligonucleotides and proceed with step 23 afterwards.

21. Purify the oligonucleotide by ethanol precipitation. Add 2 µL of 2 M $MgCl_2$ and 1 mL of cold ethanol and mix the solution. Incubate the mixture on dry ice for 40 min. Spin the mixture (15,000×g, 30 min, 4 °C). Remove the supernatant and wash with 1 µL ethanol. Spin the mixture again (15,000×g, 15 min, 4 °C) and remove the supernatant. Be careful not to disturb the precipitated pellet on the bottom.

22. Dissolve the precipitated pellet in 500 µL ddH_2O and add it to a filter unit (3 kDa, 0.5 mL). Spin the solution (14,000×g, 30 min, 4 °C) to remove residual DTT and determine the final concentration of the oligonucleotide by UV absorption at 260 nm. Use the extinction coefficient provided by the manufacturer or calculate it using the nucleic acid sequence (Internet resource: "*OligoAnalyer 3.1*[115]"). If a strong absorbance at 260 nm and a shoulder peak above 300 nm is observed repeat this step to remove excess DTT. Again, keep the flow-through to test for broken filter units. Immediately proceed to step 23.

Steps 23-26: Conjugate oligonucleotides and nanocrystals (about 8 hours)

23. Add a six-fold molar excess of purified and desalted gold nanocrystals to the reduced oligonucleotide and vortex the mixture (i.e. add 300 nmol Au particles to 50 nmol of oligonucleotides). Add 20 µL of 1 M Tris-HCl, pH 9.0 and vortex again. Incubate the solution for 2 h at room temperature.

24. Add 15 µL of 2 M ammonium acetate, pH 5.6 to stop the reaction and store the mixture on ice.

25. Purify the solution using the Dionex DNAPac 200 and anion exchange HPLC. Elute the conjugates with a salt gradient using a low salt acetate buffer (10 mM NaCl, 20 mM ammonium acetate, pH 5.6) and a high salt acetate buffer (1.5 M NaCl, 20 mM ammonium acetate, pH 5.6). Again, tune the period of the gradient according to the length of your oligonucleotide. Typically, 15-mer DNA-gold nanocrystal conjugates elute around 40% of the high salt buffer. Set the flow to 3 mL/min and monitor the absorbance at 260 nm and 360 nm. The oligonucleotides only absorb at 260 nm whereas the gold nanocrystals absorb at both wavelengths. Typically, oligonucleotides with a single gold nanocrystal elute earlier than unlabeled oligonucleotides of the same length. Free gold nanocrystals

elute very early and gold nanocrystals with multiple oligonucleotides elute later than 1:1 conjugates. For a detailed chromatogram see [83]. Use a small amount of sample to perform an initial analytical run, so that you can to make adjustments to the salt gradient if required.

26. Desalt the gold-oligonucleotide conjugates using centrifugal filter units (3 kDa, 4,000×g, 35-40 min, 4 °C) and ddH_2O. Repeat this step three times. Determine the concentration of the purified conjugate by measuring the absorption at 360 nm (0.076 µM/cm). The desalted conjugates are stable and can be stored at −20 °C for months. Typically, 12 nmol of sample can be recovered.

Steps 27-29: Preparation of final duplex conjugates for a sample quartet (about 5 hours)

27. Mix pairs of complementary single-stranded oligonucleotides in equimolar ratio and incubate DNA at room temperature or RNA at 40 °C for 30 min. A samples quartet consists of a double-labeled sample with two modified strands, two single-labeled samples with a single modified strand, and an unlabeled sample (see Fig. 4.1 and Fig. 4.4). Use unmodified oligonucleotides for the unlabeled sample or as the complementary strand for the single-labeled samples.

    Note: Use desalted oligonucleotides from step 26. In case your structure does not form in ddH_2O and room temperature only, add the required buffer and salt to the solution and perform thermal annealing.

28. Purify the annealed constructs by anion exchange HPLC using the same approach as in step 25. Use a small amount of sample to perform an initial analytical run so that you can make adjustments to the salt gradient if required. The duplex constructs typically elute later then single-stranded conjugates.

29. Collect the desired HPLC fractions and immediately desalt them using centrifugal filter units (10 kDa, 4,000×g, 15 min, 4 °C) and ddH_2O. Repeat this step three times. Determine the concentration of the purified conjugate by measuring the absorption at 360 nm (0.076 µM/cm for single-labeled samples and 0.152 µM/cm for double-labeled samples). Typically, 2-3 nmol of each final duplexed construct should be obtained. The desalted conjugates are stable and can be stored at −20 °C for months.

4.3 Basic protocol 2, Collecting X-ray scattering interferometry data

To date, XSI data has been successfully recorded at beamline 4-2 of the Stanford Synchrotron Radiation Lightsource (SSRL) [69, 82, 105–108], beamline 12-ID of the Advanced Photon Source (APS) [69, 82], the SIBLYS beamline of the Advanced Light Source (ALS) [41], and the BM29 beamline of the European Synchrotron Radiation Facility (ESRF) (see Fig. 4.8). In general, measurements can be carried out at any synchrotron with beam lines set up for SAXS measurements that meet the following requirements:

- S-range: 0.0015 – 0.12 Å^{-1} (optimal, for details see below) corresponding to a q-range 0.01 – 0.8 Å^{-1}
- X-ray energy: 9 keV-15 keV (9-11 keV is optimal). This is the tested energy range used in experiments to date, for details see below.
- Sample volumes: 16-40 µL (typical amounts used at state-of-the-art synchrotrons [19]). This amount allows for 10 independent exposures without requiring large quantities of sample.

It is important to pay attention to the definition of the magnitude of the momentum transfer vector S as two different conventions are used. In this protocol S is defined as $S = 2\sin(\theta/\lambda)$ (q is defined as $4 \cdot \pi \cdot \sin(\theta/\lambda)$), where λ is the X-ray wavelength and θ is half the total scattering angle. S is alternatively reported in units of Å^{-1} and nm^{-1}. We report S in Å^{-1} in this protocol. Set up a sample-to-detector distance that covers an S-range from 0.0015 – 0.12 Å^{-1} (for example, this corresponds to a sample-to-detector distance of 1.1 m for the Pilatus 300K detector at 11 keV on Stanford beamline 4-2). Typically, the sample and detector configuration must be arranged with beamline scientists well in advance of data collection, since it requires hardware alignment and calibration. If the beamline cannot cover the full "optimal" S-range (0.0015 – 0.12 Å^{-1}), S_{max} should not be less than 0.095 Å^{-1} as the labeled samples and the bare gold particles contribute scattering intensity up to 0.085 Å^{-1}. This is crucial for obtaining a valid interference profile by the analysis procedure described below. If the chosen beamline cannot reach this S_{max} for lower X-ray energies, one solution is to extend the S-range by selecting higher energies (for example, using 15 keV instead of 11 keV). However, X-ray energies close to gold absorption edges (L-III at 11.92 keV, L-II at 13.73 keV and L-I at 14.35 keV) should be avoided and energies below L-III are preferable to minimize X-ray fluorescence from these edges. Be aware that important details of the scattering profile can be lost in the low S-range for X-ray energies chosen too low or depleted by low signal-to-noise at X-ray energies set too high. For general protocols on SAXS sample preparation, data collection, and data analysis see [43, 116–122].

Materials

- Scattering standard sample (e.g. cytochrome c)
- Tris-HCl buffer, pH 7.4
- Sodium ascorbate
- Sodium chloride (NaCl)
- Purified gold nanocrystal sample for titration series (10 nmol)
- Purified double-labeled sample (1 nmol per buffer condition)
- Purified 2× orthogonal single-labeled samples (1 nmol per buffer condition, each)
- Purified unlabeled sample (1 nmol per buffer condition)
- Sample buffer
- MilliQ water

- UV/Vis spectrometer
- Vortex mixer
- −20 or −80 °C freezer

- Dry ice
- Microfuge
- 0.5 mL centrifugal filters (Amicon Ultra)

The protocol below is a suggested workflow for data acquisition at a synchrotron facility. It may vary based on the instrument setup at the facility.

Sample preparation and setup initialization

1. If you are not familiar with the beamline and the settings, record data for a standard, e.g. cytochrome *c* [19], and compare it to benchmark profiles (Internet resource: "*Small Angle Scattering Biological Data Bank* [123]").

2. Prepare a 10× buffer mixture containing Tris-HCl and sodium ascorbate together with the desired amount of additional salt and other components (e.g. ligands) required for the experiment. An example 10× buffer solution for near-physiological conditions is 700 mM Tris-HCl, pH 7.4, 100 mM sodium ascorbate, 1.5 M NaCl, and 10 mM $MgCl_2$. Use a Vortex mixer to ensure proper mixing of the components. It is important to use sodium ascorbate and Tris-HCl as radical scavengers in the buffer solution to capture free radicals and thus to reduce radiation damage to your sample during X-ray exposure; this allows longer total exposure times and therefore better signal-to-noise. Replace the sodium ascorbate stock solution every 3 h to ensure good scavenger capability. Cover the sodium ascorbate with aluminum foil or store it at a dark place.

3. The total exposure time depends on the flux at the synchrotron beam line used. A typical scheme, used at beam line 4-2, is to set the total exposure time to 30 s as a series of 10 independent repeats of 3 s each for data collection. Screen each trace for radiation damage, which can be detected by a gradual change in scattering intensity especially in the low *S*-range region in subsequent X-ray exposures. Exclude scattering profiles with oxidative damage determined from subsequent analysis. Furthermore, do not reuse samples that have been exposed to the X-ray beam. If the photon flux is much less than 10^{12} photons/s, extend the total exposure time.

4. If you have not done so previously (see step 29 in Preparation of gold nanocrystal-nucleic acid conjugates), determine the concentration of your sample using a UV/Vis spectrometer. The extinction coefficient of the gold nanocrystals is 0.075 μMM/cm at 360 nm.

5. Record a titration series of gold nanocrystal for every beam time as a scattering standard and to obtain the nanocrystal size distribution required for the further analysis. A typical concentration series is 200, 100, 50, and 25 μM of gold particles (include higher concentrations if they are necessary for your experiment). The shape of the scattering profile should not change with concentration, and the scattering profiles should be superimposable after normalization. Interparticle scattering should be avoided. It can be detected by a concentration-dependent change in the scattering profile at low *S*.

6. Store gold-labeled samples on dry ice or in a freezer at −20 °C until the beam line is set up for the experiments. During data acquisition, store the sample stock solutions on ice or in a fridge.

7. Thaw the required amount of sample to room temperature and vortex before the measurements.

8. Combine 10× buffer, water and the concentrated sample to achieve a final sample concentration of 30 µM (for example 0.9 nmol of sample in 30 µL). If it is not possible to prepare the buffer as a 10× stock solution (for example due to solubility limitations), or if the concentration of charge of the macromolecule is comparable to the concentration of counterions in solution at very low ionic strength, prepare the sample by buffer exchange using centrifugal filter units with a suitable molecular weight cutoff (e.g., 10 kDa Amicon, three repeats, 35 min each).

9. Spin the final sample mixture for 2 min at 10,000×g at 4 °C to sediment out any large contaminant particles. Large particles strongly perturb the scattering signal, as the forward scattering intensity of an object grows quadratically with its molecular mass.

Data recording

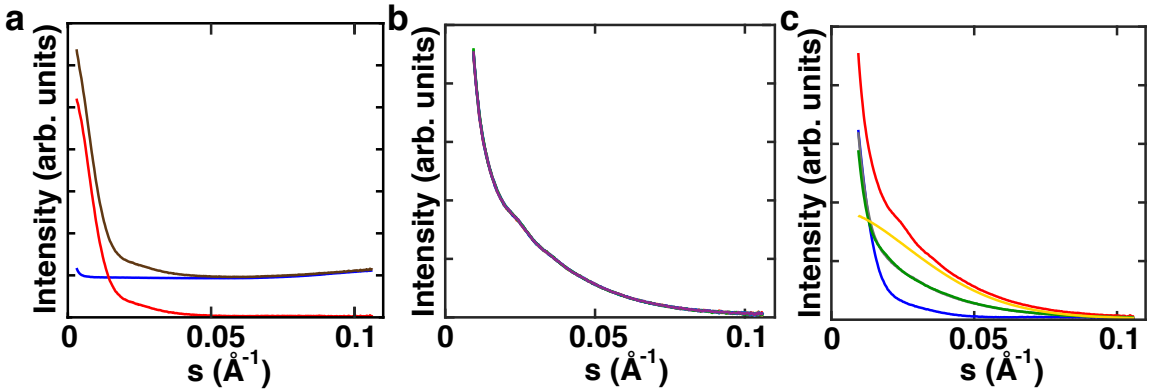

Figure 4.4: Example scattering profiles of XSI samples. (**a**) Buffer-only scattering profile used for buffer subtraction (blue), scattering profile of an unlabeled sample without (brown) and with buffer subtraction (red). (**b**) Ten individual exposures for a double-labeled sample recorded in one run. All profiles match and thus radiation damage can be excluded. These scattering profiles are buffer subtracted. (**c**) One full set of samples including one double-labeled sample (red), two orthogonal single-labeled samples (green, grey), bare gold nanocrystals (yellow) and unlabeled sample (blue). All scattering profiles are buffer subtracted.

10. Prepare five samples for data collection. The full set of samples consists of an unlabeled molecule, two complementary single-labeled samples, one double-labeled macromolecule and a buffer-only sample (Fig. 4.4). Again, the concentration of the macromolecule should be at least 30 µM in 1× buffer to provide good signal-to-noise. Measure the scattering profiles of the five samples on the same set up, in direct succession, to keep the conditions as similar as possible. Always record the buffer-only scattering profile (at least) twice, once at the beginning of the acquisition sequence and once at the end (for example:

buffer, unlabeled molecule, two complementary single-labeled samples, double-labeled macromolecule and buffer again). An automated sample changer installed at the beamline can aid data collection for such a series.

11. Repeat the 5 sample data acquisition sequence with each macromolecule construct and/or condition in your experiment (for example at varying salt concentrations, with and without ligand binding partners, etc.).

4.4 Basic protocol 3, Analyzing X-ray scattering interferometry data

The data obtained at the beamline can be processed either by applying individual scripts step-by-step or using the custom written GUI in Matlab ("*au_saxs_gui.m*", see Materials). The underlying principles are described in detail by Mathew-Fenn et al. [69] and a brief summary is as follows. After standard SAXS data processing, as outlined below, the radius distribution of the spherical gold nanocrystals is determined first, from scattering data of the unconjugated gold labels. To accomplish this, the recorded scattering profile of the gold nanocrystals is decomposed into a volume-weighted sum of scattering profiles of spheres with radii ranging from 1 to 100 Å. Using the nanocrystal synthesis protocol described above, [69, 82, 83] the nanocrystal size distribution should have a radius centered at 6-7 Å. The obtained radius distribution is then used to calculate the precise basis scattering functions $I(S, D)$, which are the scattering interference patterns for two nanocrystal separated by a fixed center to center distance D, where D is varied from 1 to 200 Å. These basis functions will be used to decompose $I_{Au-Au}(S)$, which is the experimentally-determined scattering interference pattern for the two gold nanocrystals attached to the macromolecule. Importantly, the measured scattering profile from the double-labeled macromolecule includes contributions from the macromolecule itself and from the cross-scattering terms between the gold labels and the macromolecule, in addition to I_{Au-Au}. To extract I_{Au-Au}, the scattering profiles for the quartet of samples must be summed, and the summation requires that the profile intensities are accurately scaled relative to each other. Finding the correct scaling factors, denoted c_U, c_{A+B} and c_{Buf} (see Eq. 4.2 and 4.3), is the most difficult part of the data processing. To do this, the measured scattering profiles of the double-, single-, and unlabeled constructs and of the buffer ($I_{AB}(S)$, $I_A(S)$, $I_B(S)$, $I_U(S)$ and $I_{Buf}(S)$) are first transformed into interatomic Patterson distributions $P_{AB}(D)$, $P_A(D)$, $P_B(D)$, $P_U(D)$ and $P_{Buf(}D)$ using point scatterer basis functions (see Eq. 4.1). Using the measured scattering profiles and the Patterson distributions, the scaling factors are optimized to satisfy two constraints: *i*) that the integral of the sinusoidal function $S \cdot I_{Au-Au}$ sums to zero (Eq. 4.4), and *ii*) that none of the computed interatomic distances between gold labels are negative, which should not be possible (*Eq*.5). Deviations from these constraints are summed together in a target function T (see Critical Parameters and Troubleshooting), and the scaling factors that minimize T are determined. Using the optimized scaling factors, the measured profiles are summed to obtain I_{Au-Au}.

$$I(S) = \sum_{D_{min}}^{D_{max}} P(D) \cdot \frac{\sin(2\pi DS)}{(2\pi DS)} \tag{4.1}$$

$$I_{Au-Au}(S) = I_{AB}(S) + c_U \cdot I_U(S) - c_{A+B} \cdot (I_A(S) + I_B(S)) + c_{Buf} \cdot I_{Buf}(S) \tag{4.2}$$

$$P_{Au-Au}(D) = P_{AB}(D) + c_U \cdot P_U(D) - c_{A+B} \cdot (P_A(D) + P_B(D)) + c_{Buf} \cdot P_{Buf}(D) \tag{4.3}$$

$$\sum_{S_{min}}^{S_{max}} I_{Au-Au}(S) \cdot S \approx 0 \tag{4.4}$$

$$P_{Au-Au}(D) > 0 for D \in [D_{min}; D_{max}] \tag{4.5}$$

Finally, I_{Au-Au} is decomposed into a sum of the $I(S,D)$ basis functions using a maximum entropy algorithm, resulting in a center-to-center distance distribution between the two gold nanocrystals of the double-labeled sample (Eq. 4.1). Alternatively, the I_{Au-Au} decomposition can be performed using non-negative least squares algorithms that are available in most scientific programming languages. Some beamlines, such as the beamline 4-2 at SSRL and beamline BM29 at ESRF, provide beamline software packages that perform radial averaging and buffer subtraction of scattering profiles, which allows for immediate detection of radiation damage or other technical problems.

Materials

Data set including scattering profiles from:

- Gold nanocrystal sample
- Double-labeled sample
- Orthogonal single-labeled samples (two individual scattering profiles)
- Unlabeled sample
- Sample buffer

- Computer (Minimum requirements: Any Intel or AMD x86-64 processor, 2.5 GB Disk space, 2 GB RAM)
- MATLAB license (GUI support from version 2015b guaranteed)
- Au-SAXS GUI (`https://gitlab.physik.uni-muenchen.de/Jan.Lipfert/AuSAXSGUI.git`)
- Example files (`https://gitlab.physik.uni-muenchen.de/Jan.Lipfert/AuSAXSGUI.git`)

A step-by-step guide on how to obtain the scattering interference pattern I_{Au-Au} including an example set of data (see exemplary files) is given below.

Data preparation

1. If it has not already been done automatically by the beamline software, reduce the 2D scattering matrix into a one dimensional scattering profile by radial averaging. The

output should be a matrix with three columns for the scattering momentum transfer vector S, the corresponding scattering intensity, and the variance/standard deviation in scattering intensity at different radial positions.

Note: If the incident X-ray beam is polarized, simple radial averaging cannot be performed. See [66, 124] for further instruction on how to process the raw data into one dimensional scattering profiles.

2. Before starting the custom written *au_saxs_gui.m* GUI and loading the data, rename the data files '*_i.dat' where * is any name for the sample and i is the i^{th} exposure, i.e., ranging from 01 to 10 for 10 exposures per molecule ('AB_01.dat', 'AB_02.dat', 'AB_03.dat', ..., for double-labeled samples). Structure the data files so that the scattering momentum vector occupies the first column, the corresponding recorded scattering intensity occupies the second column and the variance/standard deviation occupies the third column, and the column entries are separated by a single blank space (see example files for comparison).

3. Initialize the GUI by executing the 'au_saxs_gui.m' script. For proper execution the files 'au_saxs_gui.m', 'au_saxs_gui.fig' and the folder 'subroutines' have to be stored in the same directory to allow the main script to find the required subroutines.

Data initialization

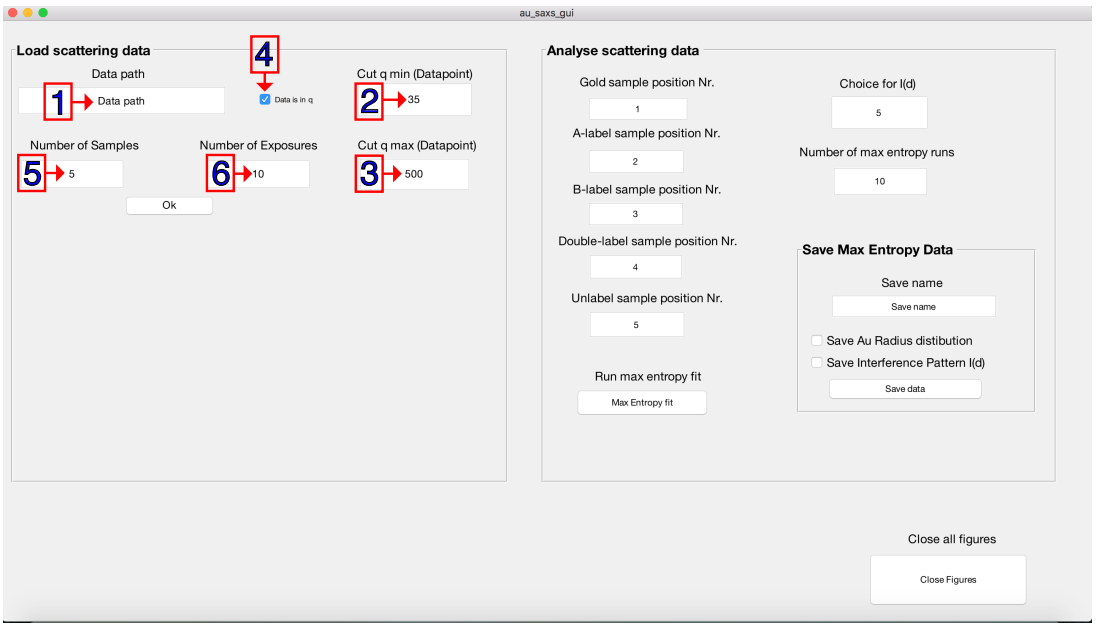

Figure 4.5: GUI to analyze XSI data. Default values can be modified to adjust the data analysis. The GUI contains panels to specify the following: (1) storage path of the data; (2) adjust q_{min} 2); (3) adjust q_{max}; (4) specify the momentum transfer convention according to the scattering data; (5) specify the number of samples; and (6) specify the number of exposures per sample.

4. Enter the full path into the field 'Data path' (Fig. 4.5.1).

5. Optional: Manipulate the scattering momentum by setting the lower (q_{min}) (Fig. 4.5.2) and the upper (q_{max}) (Fig. 4.5.3) limit for the scattering angle. The default input is the 35th data point of the initial data up to data point 500; however, this strongly depends on the settings of the beamline and the type of sample.

6. Optional: Untick the 'Data is in q' box (Fig. 4.5.4) to switch the scattering momentum vector to S ($2\sin(\theta/\lambda)$ in Å^{-1}).

   Note: The scattering momentum vector is set to q per default ($4\pi sin(\theta/\lambda)$, in Å^{-1}) since q is the common output format at synchrotrons.

7. Optional: Modify the number of samples (Fig. 4.5.5) if required. (Default is 5 for one full set of samples.)

8. Optional: Change the number of exposures (Fig. 4.5.6) the number of files read per sample. (Default is 10 as described in the protocol.)

9. Press 'Ok' to initialize the script. After doing so new buttons and a table will appear on the left side of the window (see Fig. 4.6).

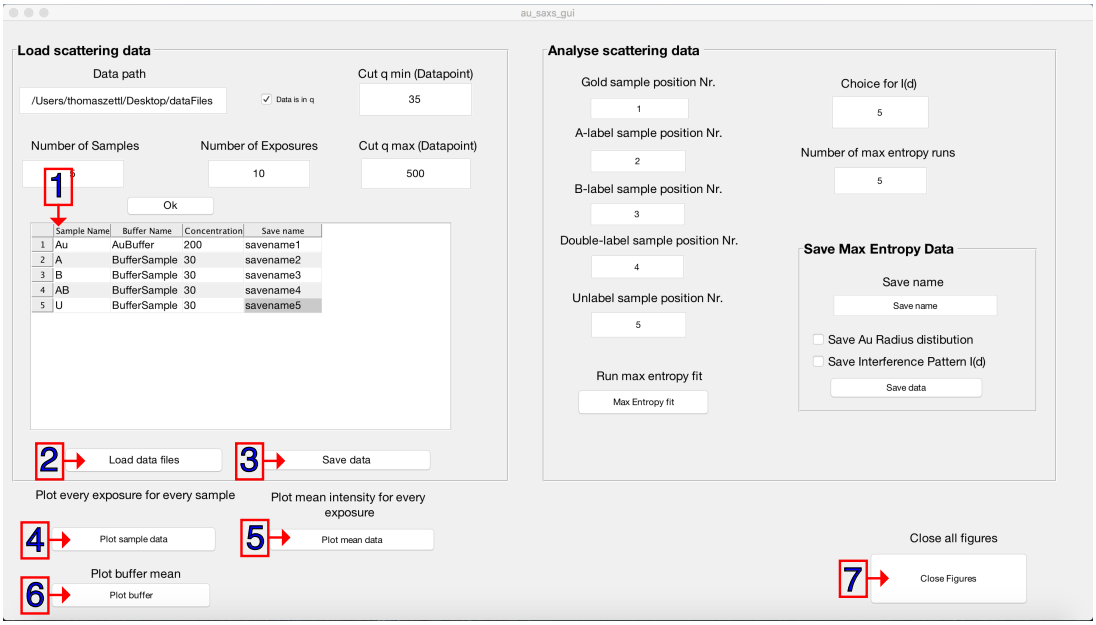

Figure 4.6: GUI to analyze XSI data. Sample and corresponding buffer filenames, sample concentration and save as filename (optional) are entered for a full set of data (1). Initial files are loaded (2) and can be plotted (4-6) (optional). Unified and truncated scattering profiles can be saved (3).

10. Enter all sample names in the first column of the table (see Fig. 4.6).

11. Enter the corresponding buffer in the second column.

12. Enter the determined sample concentration (in µM) in the third column.

13. Optional: Enter save as filenames for all samples in the fourth column (see Fig. 4.6.1).

14. Load the data files by pressing the ‘Load data files’ button (Fig. 4.6.2). After successfully doing so, a window indicating ‘All data was successfully imported!’ will be displayed. In case a file could not be found or a wrong sample name was entered a warning will be generated noting to the incorrect position. If the concentration was not entered properly a similar error message will be produced.

Data testing

15. Optional: Inspect the scattering profiles for all samples either by plotting all exposures per sample as individual traces into one figure per sample (Fig. 4.6.4) or by plotting the averaged profile over all exposures (Fig. 4.6.5).

16. Optional: Overlay these plots by pressing the ‘Plot sample data’ first and the ‘Plot mean data’ second.

17. Optional: Plot the averaged buffer for each sample using the ‘Plot buffer’ button (Fig. 4.6.6).

18. Optional: In case too many MATLAB figures are open, close all except for the main GUI by pressing the ‘Close Figures’ button (Fig. 4.6.7).

19. Optional: Save the scattering data for all loaded samples as ‘YOURFILENAMEHERE_scattering_data.mat’ files using ‘Save data’ (Fig. 4.6.3).

Maximum entropy fitting

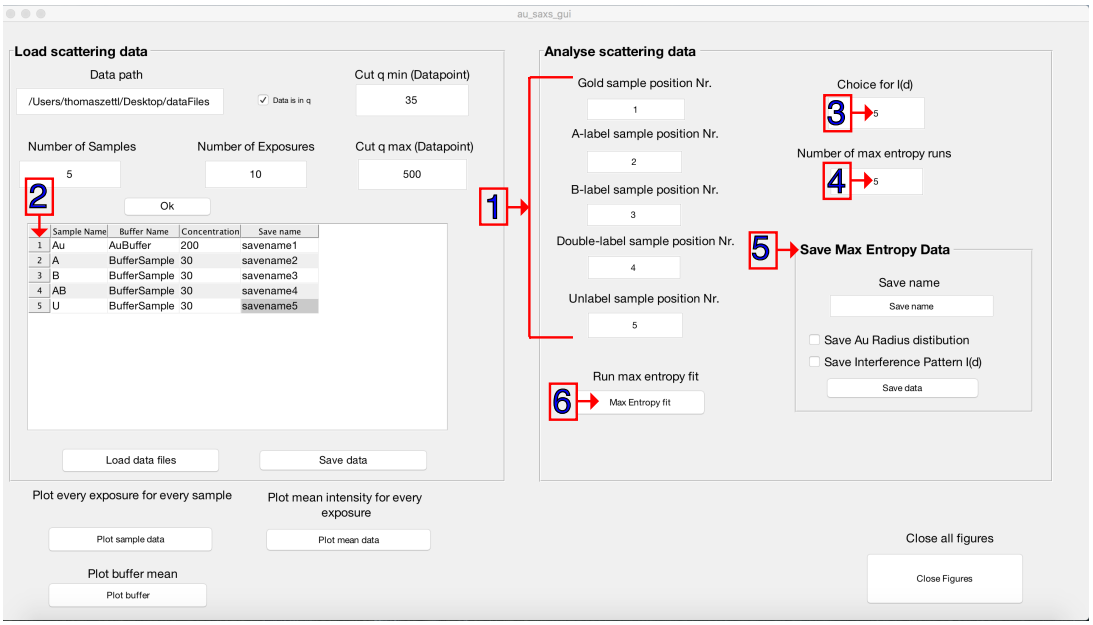

Figure 4.7: GUI to analyze XSI data. Panels to set the sample type (1) according to the position in the sample loading table (2). The choice of minimization function (3) and the number of individual maximum entropy fitting runs (4) can be specified prior to starting the routine (5). Final distance distributions (Au radius distribution and interference pattern I(S), optional) can be saved (6).

20. Set the options by specifying the sample positions according to the row number in the table (see Fig. 4.7.1), i.e., Gold sample position = 1, A-label sample position = 2, B-label sample position = 3, Double-label sample position = 4 and Unlabeled sample position = 5.

21. Optional: Change the number of runs for the maximum entropy fit (default is 10, Fig. 4.7.3). The output is one high-resolution distance distribution per run. Average the distributions over all runs to obtain the final distribution. Lower the number to shorten the required computational time for test purposes.

22. Optional: Change the minimization function option to extract the gold-gold interference pattern I_{Au-Au} (default is 5 ranging from 1 to 7, see Critical Parameters and Troubleshooting for more details, Fig. 4.7.2).

23. Start the maximum entropy fit by pressing 'Max Entropy fit' (Fig. 4.7.4). A progress bar in the lower left corner of the main GUI will display the progress and vanish as soon as the calculations are finished. Three new figures showing the gold nanocrystal radius distribution, the gold-gold scattering interference signal I_{Au-Au}, and the final distance distribution determined via the maximum entropy fit will appear (Fig. 4.8).

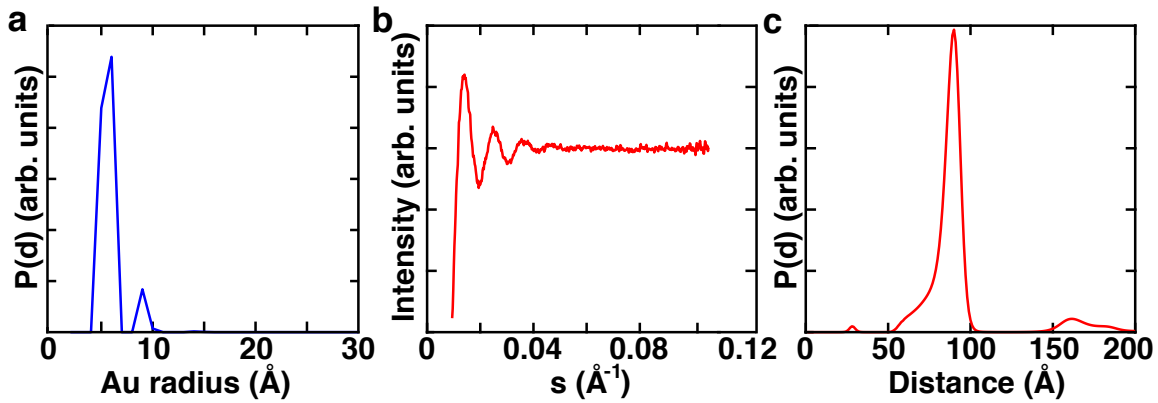

Figure 4.8: Maximum entropy analysis of XSI data. (**a**) Radius distribution of gold nanocrystals used to generate basis profiles $I(S, D)$. (**b**) Calculated interference pattern for the example data set and (**c**) final distance distribution averaged over 10 maximum entropy fitting runs.

24. Enter a save as filename (e.g. 'YOURSAMPLENAME') for the distance distribution (Save Max Entropy Data, right side, Fig. 4.7.5). This file has one column per maximum entropy run (10 as default) and a 1 Å spaced distance distribution ranging from 1 Å up to 200 Å.

25. After setting the name, press 'Save data' (Save Max Entropy Data, right side) a file in the format of 'YOURSAMPLENAME_Distance_Distribution.mat' will be saved in your current folder.

    Optional: Save the gold nanocrystal radius distribution and/or the Au-Au interference pattern I_{Au-Au} checking the individual boxes.

4.5 Reagents and Solutions

FPLC cleaning solution:
100 mM DTT, 20 mM Tris-HCl, pH 8.0 (prepare fresh)

Size exclusion running buffer:
150 mM ammonium acetate, pH 5.6 (prepare fresh)

Ammonium acetate stock buffer:
2 M ammonium acetate, pH 5.6 (store up to one month at 4 °C)

Low salt borate buffer:
10 mM NaCl, 20 mM Na-borate, pH 7.8 (store up to one month at room temperature)

High salt borate buffer:
1.5 M NaCl, 20 mM Na-borate, pH 7.8 (store up to one month at room temperature)

Low salt ammonium acetate buffer:
10 mM NaCl, 20 mM ammonium acetate, pH 5.6 (store up to one month at room temperature)

High salt ammonium acetate buffer:
1.5 M NaCl, 20 mM ammonium acetate, pH 5.6 (store up to one month at room temperature)

4.6 Background Information

Basic principle

XSI measures the interference between X-rays scattered by two gold nanocrystals attached to a macromolecule or macromolecular complex. In addition to the desired nanocrystal-nanocrystal interference pattern, the recorded scattering profile includes signals from scattering interference between pairs of atoms in the macromolecule, between pairs of atoms in a single gold nanocrystal, and between pairs of atoms in the macromolecule and the nanocrystal. It is critical to isolate the nanocrystal-nanocrystal interference pattern from the other signals, since it alone contains direct information about the distribution of center-to-center distances between the nanocrystal probes. This distance distribution is obtained by Fourier transformation of the nanocrystal-nanocrystal interference pattern. A limitation of XSI is the time-consuming sample preparation, which only allows low throughput. The strength of XSI is that it provides accurate ensemble distance distributions even for dynamic and rapidly interconverting macromolecules, with precise external distance calibration from the wavelength of the incident X ray radiation. These absolute distances can be directly compared to high-resolution atomic models from NMR spectroscopy, X-ray crystallography, electron microscopy or molecular simulations. XSI is very complementary to spectroscopic rulers such as FRET, and combined application of the two types of ruler presents interesting possibilities.

Application of XSI to nucleic acids and nucleic acid-protein complexes

Over the last decade, XSI has been successfully applied to dsDNA [69, 82, 105], nucleic acid two-way junctions [106, 108] and nucleic acid-protein complexes [41, 107]. It has been used

to characterize the stretching [82], twisting, and bending elasticity [105] of short DNA helices. Recently, DNAs and RNAs containing bulge sequences have been studied, revealing a complex multistate behavior that responds to the solution conditions [106, 108]. In addition, an RNA-protein complex has been investigated [107]. The use of metal clusters as structural probes makes XSI scalable. Clusters of a single or few metal atoms are suitable for smaller systems [81, 125], while clusters containing thousands of atoms can be applied to large complexes [41]. While small labels are desirable for accurate position determination and to minimize the perturbation of the molecular structures of interest, for large macromolecules noise from the macromolecule signal can prevent accurate Au-Au distance determination and larger gold nanocrystals are needed.

A variant of the XSI interference technique has recently been demonstrated that uses anomalous SAXS to directly extract the I_{Au-Au} interference pattern from double-labeled samples without the need to record any single- or unlabeled samples [2]. This approach will be in particular relevant for samples where it is difficult to prepare matching single labeled constructs.

Future directions

The future challenges and possible applications of XSI are numerous. Using the labeling and measuring protocols described here, the conformational ensembles of diverse nucleic acid motifs and nucleic acid protein complexes can be determined, including their dependence on environmental conditions. The application of XSI to nanocrystal-labeled proteins will provide a new window into protein conformational ensembles, especially for folding intermediates and intrinsically disordered proteins (IDPs). Finally, the application of next-generation free electron lasers with very high brilliance could allow measurement of correlated and time resolved distances between multiple sites in a macromolecule, with a time resolution of tens of femtoseconds [86, 89, 126–132].

Critical Parameters and Troubleshooting

A key aspect for successful determination of intramolecular distance distributions using XSI is sample preparation and sample quality. The samples have to be highly homogeneous (e.g., no free gold nanocrystals in solution) and free of degradation. Furthermore, buffer matching should be as precise as possible. Given that XSI requires very pure samples it is crucial to monitor and adjust the sample quality in advance. To assess sample quality, non-denaturing polyacrylamide gel electrophoresis (PAGE) [133] or ion exchange chromatography [134] can be used. In addition, we advise the user to perform additional experiments (e.g., UV melting curve analysis and circular dichroism spectroscopy, see [69]) to test whether the attached gold nanocrystals perturb the macromolecular structure and, ideally, to establish the absence of such perturbations. In addition, G/C base pairs at the blunt ends of nucleic acids are recommended to avoid fraying of nucleic acid helices.

Another requirement of the XSI technique is high signal-to-noise SAXS measurements of the set of double-, single-, and unlabeled samples. We recommend testing the beam line set up with readily available and characterized samples, such as cytochrome *c*, bovine serum albumin (BSA), or unlabeled double-stranded DNA (Internet resource: *"Small Angle Scattering Biological Data Bank"*[123]), prior to an XSI experiment. The measurements of the scattering

standard samples should give the expected scattering profiles and radii of gyration. Such control measurements should be repeated every time the SAXS setup is reconfigured and can help to detect misalignment of the X-ray beam, problems with parasitic scattering, and issues in detector calibration and conversion of the detector image to the 1D scattering profile.

Sometimes, it is not possible to obtain a high-quality distance distribution even when the sample preparation and recording are performed correctly. Potential issues can be the scattering intensity ratio between the molecule and the gold nanocrystals resulting in poor signal-to-noise for larger nucleic acid structures or highly flexible molecules. Additionally, larger metal clusters should be considered in cases of insufficient signal-to-noise.

Some additional factors that can affect the data analysis and the final distance distribution and how to address them are listed below:

- Parasitic X-ray scattering in the low S-range as well as high noise levels in the high S-range can occur and influence the normalization and calculation of the interference pattern.

  Solution: We recommend measuring scattering standards to test and optimize the X-ray set up before running XSI samples. In the post-processing of recorded data sets, one should vary S_{min} and/or S_{max} multiple times and test the impact on the resulting distance distribution.

- The scattering profile can be distorted by various effects such as radiation damage or bubbles passing through the X-ray beam.

  Solution: Carefully compare scattering profiles from subsequent exposures of the same sample and exclude all traces differing from the majority.

- Choice of minimization function T to extract to gold-gold scattering interference Patterson I_{Au-Au} where S is the scattering momentum vector, I_U the scattering intensity vector of the unlabeled sample and P_{Au-Au} the label-label interatomic radial Patterson. The gold-gold scattering interference Patterson I_{Au-Au} looks like a sinc (i.e. a decaying oscillation) function (see Fig. 4.8b). If that is not the case redo the analysis with a different choice for T. As guidance and for comparison, multiple interference pattern of static and dynamic systems can be found in the published literature [69, 82, 83, 105–108] for comparison.

  Choice 1: Original function described by Mathew-Fenn, Das and coworkers [69]

$$T = \frac{\sum_S [I_{Au-Au} \cdot S^2]^2}{\sum_S [I_U \cdot S^2]^2} + \frac{\sum_{P_{Au-Au}(D)<0} P^2_{Au-Au}(D)}{\sum_D P^2_U(D)} \tag{4.6}$$

  Minimization function similar to Kratky analysis dividing out the decay of the scattering intensity by weighting the interference pattern I_{Au-Au} by S^2.

  Choice 2:

$$T = \frac{\sum_S [I_{Au-Au} \cdot S]^2}{\sum_S [I_U \cdot S^2]^2} + \frac{\sum_{P_{Au-Au}(D)<0} P^2_{Au-Au}(D)}{\sum_D P^2_U(D)} \tag{4.7}$$

  Au-Au interference pattern is only weighted by S, thus I_{Au-Au} minimization is dominated by values in the low S-region.

Choice 3:

$$T = \frac{\sum_S [I_{Au-Au} \cdot S]^2 \cdot S}{\sum_S [I_U \cdot S^2]^2} + \frac{\sum_{P_{Au-Au}(D)<0} P^2_{Au-Au}(D)}{\sum_D P^2_U(D)} \tag{4.8}$$

The low S-region of I_{Au-Au} is weighted more compared to choice 1 4.6, but less than choice 2 4.7 and the high S-region is weighted less compared to choice 1 4.6 and more than choice 2 4.7.

Choice 4:

$$S_{min} = Min(S) \quad \text{for} \quad S_i \in S \geq 0.06 \text{ Å}^{-1}$$
$$S_2 = \frac{S_i}{S_{min}} \quad \text{for} \quad S_i \in S < 0.06 \text{ Å}^{-1}$$
$$S_2 = 1 \quad \text{for} \quad S_i \in S \geq 0.06 \text{ Å}^{-1}$$

$$T = \frac{\sum_S [I_{Au-Au} \cdot S]^2 \cdot S_2}{\sum_S [I_U \cdot S^2]^2} + \frac{\sum_{P_{Au-Au}(D)<0} P^2_{Au-Au}(D)}{\sum_D P^2_U(D)} \tag{4.9}$$

Very similar to choice 3 4.8 in the high S-range, however the high S-range is weighted slightly less.

Choice 5:

$$S_{min,1} = Min(S_i) \quad \text{for} \quad S_i \in S < 0.06 \text{ Å}^{-1}$$
$$S_{min,2} = Min(S_i) \quad \text{for} \quad S_i \in S \geq 0.06 \text{ Å}^{-1}$$
$$S_2 = \frac{S_i}{S_{min,2}} \quad \text{for} \quad S_i \in S < 0.06 \text{ Å}^{-1}$$
$$S_2 = \frac{S_{min,1}}{S_{min,2}} \quad \text{for} \quad S_i \in S \geq 0.06 \text{ Å}^{-1}$$

$$T = \frac{\sum_{S_2} [I_{Au-Au} \cdot S]^2 \cdot S_2}{\sum_S [I_U \cdot S^2]^2} + \frac{\sum_{P_{Au-Au}(D)<0} P^2_{Au-Au}(D)}{\sum_D P^2_U(D)} \tag{4.10}$$

Very similar to choice 3 4.8 in the high S-range, however the high S-range does contribute very little.

Choice 6:

$$T = \frac{\sum_{P_{Au-Au}(D)<0} P^2_{Au-Au}(D)}{\sum_D P^2_U(D)} \tag{4.11}$$

Only $P_{Au-Au}(D)$ is minimized such that it does not include negative values, however the oscillations in I_{Au-Au} are not minimized.

Choice 7:

$$T = \frac{\sum_{S>0.06 \text{ Å}^{-1}} [I_{Au-Au} \cdot S]^2}{\sum_S [I_U \cdot S^2]^2} \tag{4.12}$$

Only the high S-region of I_{Au-Au} is minimized similar to a base line correction. $P_{Au-Au}(D)$ is allowed to include negative values.

Anticipated Results

The protocols presented here guide experimenters to synthesize gold nanocrystals, prepare and purify gold-labeled nucleic acid constructs, measure a complete set of SAXS profiles for the labeled samples, and to extract intra-molecular distance distributions from the data. Following the steps presented in Section 4.2, the following yields are typically achieved: the gold nanocrystal synthesis yields 3-10 µmol after final purification and desalting; the gold nanocrystal attachment protocol results in a yield of 2-3 nmol of each labeled and purified duplexed construct for XSI measurements.

Section 4.3 and Section 4.4 describe a reliable procedure to obtain SAXS scattering profiles and a step-by-step guide on how to use the provided GUI to extract high-resolution distance distributions on an absolute scale (Fig. 4.8c). To study more complex macromolecules that can undergo conformational changes or to disentangle complex geometries and motions, multiple label pairs attached to different positions can be prepared and analyzed to increase the information content.

Time Considerations

The sample preparation should be performed about one month in advance. The final purification (steps 28 and 29 in Section 4.2) should be carried out shortly before data collection to ensure the highest sample quality. The overall data acquisition time varies depending on the synchrotron facility. Typically, data collection for several samples can take up to 1 day. The run time of the custom-written Matlab routine to compute distance distributions that is provided with this protocol is on the order of minutes on a standard personal computer. The total time required for data analysis, validation, and interpretation is variable and can take several days.

Chapter 5

Absolute Intramolecular Distance Measurements with Ångström Resolution Using Anomalous Small-Angle X-ray Scattering

Summary

Accurate determination of molecular distances is fundamental to understanding the structure, dynamics, and conformational ensembles of biological macromolecules. Here we present a method to determine the full distance distribution between small (~7 Å radius) gold labels attached to macromolecules with very high-precision (≤1 Å) and on an absolute distance scale. Our method uses anomalous small-angle X-ray scattering close to a gold absorption edge to separate the gold-gold interference pattern from other scattering contributions. Results for 10-30 bp DNA constructs achieve excellent signal-to-noise and are in good agreement with previous results obtained by single-energy SAXS measurements without requiring the preparation and measurement of single labeled and unlabeled samples. The use of small gold labels in combination with ASAXS read out provides an attractive approach to determining molecular distance distributions that will be applicable to a broad range of macromolecular systems.

5.1 Introduction

Measurements of molecular distances are key to dissecting the structure, dynamics, and functions of biological macromolecules. While (Förster) fluorescence energy transfer (FRET) and electron (EPR) or nuclear magnetic resonance (NMR) techniques have provided invaluable de-

This Chapter was published by Zettl *et al.* [2] in *Nano Letters* and adapted with permission from the American Chemical Society. I analyzed the data and contributed to writing the manuscript with help of the other authors.

tails by measuring intramolecular distances, they suffer from a limited range (typically <10 nm) and difficulties in converting the measured signal into absolute distances or, better yet, complete distance distributions [69, 82, 135–137]. SAXS measurements employing gold nanoclusters as labels attached to DNA molecules have demonstrated their ability to provide information about the entire label-label distance distribution for a considerable range of distances ranging from 5 up to 40 nm [41, 69, 82, 105, 106, 138]. The approach has provided a detailed view of DNA structure and flexibility [82, 105, 106], revealed conformational changes of DNA upon protein binding [41], and probed the conformational landscape of a complex RNA motif in response to solution conditions and protein binding [107].

The distance distributions are obtained by inverting the gold label-label scattering interference term. According to the Debye formula this interference pattern is given by

$$I_{Au-Au}(s) = \sum_{d=0}^{d_{max}} P(d)|f_{Au}(s)|^2 \frac{\sin(2\pi sd)}{(2\pi sd)} \tag{5.1}$$

where $f_{Au}(s)$ denotes the scattering factor for a gold nanocrystal and s is the magnitude of the momentum transfer $s = 2\sin(\theta)/\lambda$ with λ as the X-ray wavelength and $2 \cdot \theta$ being the total scattering angle; d is the distance between the gold nanocrystals and $P(d)$ is the center-of-mass distance distribution. Unfortunately, the scattering from a double-labeled macromolecule does not only contain the label-label interference term (Eq. 5.1) but also scattering contributions from the gold label and macromolecule alone as well as a gold-macromolecule interference term. So far, two approaches have been employed to separate the label-label interference term from the other contributions (intragold label, gold label-macromolecule, and intramacromolecule) to the measured scattering profile. In a first approach, each of the single-labeled and the unlabeled samples are measured in addition to the double-labeled macromolecule; from addition and subtraction of the appropriate profiles, the interference term can be determined [69, 82].

While this approach has provided unprecedented insights into the conformational ensembles of nucleic acids [82, 105–107], it requires preparation and measurements of several differently labeled samples for each measured distance distribution. Preparing separate samples of the single- and double-labeled macromolecule can be challenging or impossible if the sample cannot be assembled reliably from individually labeled components. In addition, the differencing procedure requires chemically nonequivalent molecules to adopt the same conformational ensembles and the concentrations or scattering intensities of the various samples need to be carefully matched to achieve the desired separation of scattering terms. A second approach relies on using relatively large (~50 Å diameter) gold particles and neglecting the DNA and gold-DNA scattering terms [41, 138]. A drawback of this approach is the large size of the labels, which might perturb the conformational ensemble and limit its resolution. In addition, neglecting the macromolecule scattering as well as the gold-macromolecule term is problematic or unfeasible for large and strongly scattering macromolecules.

5.2 Results

Here we present an alternative strategy to separating the gold label-label interference term and instead determine the intramolecular distance distribution based on the physics of anomalous small-angle X-ray scattering [55] (ASAXS). Our approach is based on recording scattering

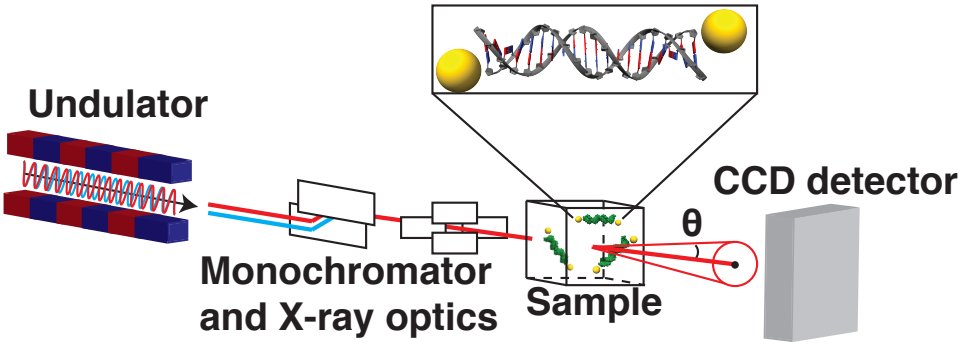

Figure 5.1: Schematic of the ASAXS measurement on double-labeled DNA molecules. The incident beam is shown along with the monochromator to select particular X-ray energies. Double goldlabeled DNA molecules are placed into the X-ray beam in a sample cell. The direct beam is blocked by a beam stop and scattered photons are detected using a CCD detector. The inset shows a 20 bp DNA molecule with two gold labels; the linkers and gold functionalization are omitted for clarity.

profiles of the double-labeled sample at different X-ray energies, scanning through a gold absorption edge. Close to an absorption edge, atomic scattering factors change rapidly with X-ray energy and take on the complex form

$$f(E) = f_0 + f'(E) + if''(E) \tag{5.2}$$

Away from the edge, the anomalous terms f' and f'' are negligible and the energy independent term f_0 dominates [55]. Because the gold absorption edges are well separated from the elements that make up biological macromolecules (C, H, N, O, P), tuning through a gold absorption edge provides a way to significantly alter the gold scattering compared to the scattering contribution from the macromolecule. The dependence on X-ray energy permits, therefore, separation of the gold-gold scattering terms from macromolecule and gold-macromolecule terms [139]. We demonstrate that our ASAXS approach enables the accurate determination of gold label-label distance distributions from measurements of double-stranded DNA constructs double-labeled with small gold clusters without the need to prepare and measure multiple molecular constructs or to use large gold labels.

We employed 10, 20, and 30 bp double-stranded DNA constructs double-labeled with 7 Å radius thio-glucose functionalized gold particles attached via thiol-chemistry as described previously [69, 82, 114] (see Supporting Information 5.4). ASAXS data were recorded at beamline 12-ID of the Advanced Photon Source [19, 140] (Fig. 5.1 and Supporting Information 5.4). Prior to measurements of DNA constructs, the X-ray energy was calibrated by inserting a thin (50 µm) gold foil into the X-ray beam and measuring the incident and transmitted X-ray intensity as a function of energy (Fig. 5.2). The data clearly indicate the position of the gold L-III edge at ~11.9 keV and allow us to reference all measurements to the tabulated absorption data (`http://henke.lbl.gov/optical_constants/`; see Supporting Information 5.4). Similar measurements using the thioglucose passivated gold nanocrystals in solution instead of the gold foil confirm that the absorption edge of the gold clusters is identical within experimental error to bulk gold (Fig. 5.2).

Experimental scattering profiles for double gold-labeled DNA constructs recorded at different energies ranging from 200 eV below to 50 eV above the gold L-III edge exhibit clear and sys-

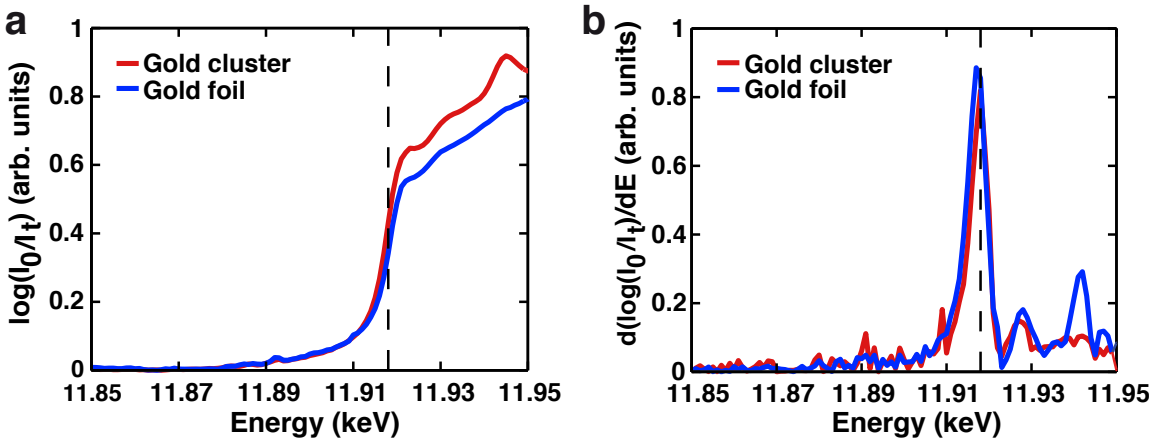

Figure 5.2: Energy calibration of the gold L-III absorption edge. (**a**) Normalized logarithm of the incident X-ray intensity divided by the transmitted intensity for a gold foil (blue) and for gold nanoparticles (red) in solution. The data show a sharp increase at the absorption edge. (**b**) Numerical derivative of the data in (a), revealing the position of the edge as a maximum in the derivative. Data are referenced (see Supporting Information 5.4) to the tabulated value of the gold L-III absorption edge at 11.919 keV (shown as dashed vertical lines).

tematic changes with X-ray energy (Fig. 5.3a and Figs. 5.5 to 5.7; see Supporting Information 5.4 for details of the experimental and normalization procedures).

To extract the gold label-gold label interference pattern from the scattering profiles recorded at different energies, we used a matrix inversion approach [139] based on the following relation (for a more detailed visualization see Supporting Fig. 5.8)

$$\begin{pmatrix} I_1(s) \\ I_2(s) \\ \vdots \\ I_{N_E}(s) \end{pmatrix} = \begin{pmatrix} a_1(s) & b_1(s) & c_1(s) \\ a_2(s) & b_2(s) & c_2(s) \\ \vdots & \vdots & \vdots \\ a_{N_E}(s) & b_{N_E}(s) & c_{N_E}(s) \end{pmatrix} \begin{pmatrix} G_{Au-Au}(s) \\ G_{Au-mol}(s) \\ G_{mol-mol}(s) \end{pmatrix} \tag{5.3}$$

$$I = TG \tag{5.4}$$

Here the $I_i(s)$ are the scattering profiles recorded at energies i and the matrix I has dimensions $(N_E \cdot N_s) \times 1$ (where N_E is the number of energies and N_s is the number of s channels). The $(N_E \cdot N_s) \times (3N_s)$ matrix T comprises precomputed matrices $a_i(s)$, $b_i(s)$, and $c_i(s)$ that are all $N_s \times N_s$ diagonal square matrices containing the energy-dependent label-label pair scattering factors, label-molecule pair scattering factors, and intramolecule pair scattering factors, respectively (see Supporting Information 5.4). The s-dependence of the $a_i(s)$ is given by the squared scattering factors for the gold labels that are approximated as spheres with a radius of 7 Å; similarly, the s-dependence of the $b_i(s)$ is given by the (non-squared) scattering factor for the gold labels (Supplementary Fig. 5.8b,c). The $c_i(s)$ have no explicit dependence on s, that is, they are constant for different values of s (Supplementary Fig. 5.8d). Importantly, this means that no assumption about the shape of the molecule scattering term is required for our analysis. The energy dependence of the $a_i(s)$, $b_i(s)$, and $c_i(s)$ is given by the energy dependence of the atomic scattering factors for gold and for the atoms in the macromolecule, respectively. We note that this means that we have assumed knowledge of the average element composition of the labeled macromolecule, as is typically the case. However, given the very

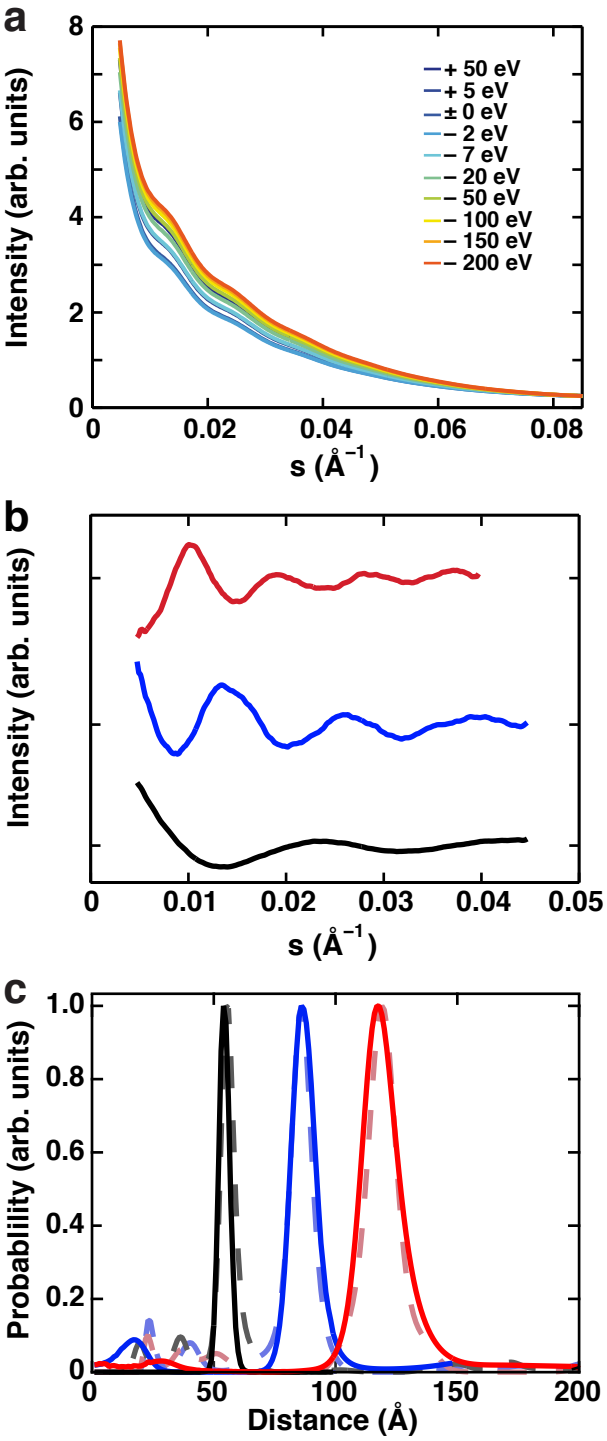

Figure 5.3: ASAXS scattering data for double-labeled gold samples and distance distributions. (**a**) Scattering intensities as a function of momentum transfer s for the 20 bp double-labeled DNA construct at 10 different energies. Energies in the figure legend are relative to the gold L-III edge. (**b**) Gold label-gold label interference patterns for 10 bp (black), 20 bp (blue), and 30 bp (red) DNA constructs obtained as described in the main text. Profiles are vertically offset for clarity. (**c**) Gold label-gold label distance distributions computed from the data in (b) (solid lines; same color code) by regularized Fourier transformation (see Supporting Information 5.4). For comparison, the distance distributions of the same samples obtained in [82] are shown as dashed lines.

minor energy dependence of the atomic scattering factors for the elements that make up biological macromolecules in the energy range used (Supplementary Fig. 5.8d, right panel), the energy dependence of the molecular term could be neglected without loss of resolution.

Finally, G represents the $(3N_s) \times 1$ vector of partial structure factors corresponding to the gold label-gold label (G_{Au-Au}), gold label-molecule (G_{Au-mol}) and molecule-molecule ($G_{mol-mol}$) terms. The vector of partial scattering factors G is obtained by (least-squares) matrix inversion of Eq. 5.4 as $G = T^{-1}I$. The gold label-label structure factor $G_{Au-Au}(s)$ was further processed [139] by subtracting a constant offset such that the mean level of oscillation approaches zero and by truncating at high s-values as the signal-to-noise ratio decreases with increasing s. We note that in principle scattering profiles at 3 different energies would be sufficient to determine G; however, using data at more (10 in our case) energies overdetermines the matrix equation (Eq. 5.4) and improves the signal [88, 141, 142]. It is difficult to give general guidelines on how many energies are required, as the results not only depend on the number of energies included, but also on the signal-to-noise of the measurements and on the positioning of the energies relative to the absorption edge. Nonetheless, model calculations on truncated data sets suggest that for our system going from 10 to 6 appropriately chosen energies leads to a loss of signal but still enables computation of meaningful gold-gold partial structure factors (Supplementary Fig. 5.9). Further reducing the number of energies or incorrectly selecting the position of the energies relative to the edge leads to a further reduction or even complete loss of signal (Supplementary Fig. 5.9). Our current choice of 10 energies is a compromise of balancing the need to avoid radiation damage (which would be an issue if significantly more exposures would be recorded for the same sample) while achieving a good signal-to-noise ratio for the computed gold-gold partial structure factors.

The gold label-gold label structure factors $G_{Au-Au}(s)$ show characteristic oscillation patterns (Fig. 5.3b) and contain information about the label-label distance distributions. The distance distributions are obtained from the $G_{Au-Au}(s)$ by regularized Fourier transformation using a maximum entropy algorithm (Supporting Information 5.4). The distance distribution $P(d)$ for the different DNA lengths all exhibit prominent approximately Gaussian peaks (Fig. 5.3c, solid lines). The center positions of the main peaks increase with increasing DNA length (Figs. 5.3c and 5.4a and Table 5.1) and are well fit by a model that takes into account the off-center attachment of the labels (Supporting Information 5.4) yielding a rise-per-base pair of 3.23 ± 0.1 Å (Fig. 5.4a). The mean positions of the peaks determined in independent repeat measurements are within $\lesssim$1 Å (Table 5.1), highlighting the exquisite precision of the method.

The variance of the main peaks increases rapidly with DNA length and the dependence of the variance of the label-label distances on DNA length is well described (χ^2 = 0.61) by a quadratic dependence with a constant offset (Fig. 5.4b, solid line). The constant offset was fit to be 1.5 Å^2 and prefactor of the quadratic term to be 0.063 Å^2/bp^2; this observation is consistent with a model that assumes cooperative stretching of the helix [82, 143] where each base pair step contributes 0.25 Å to the standard deviation. Alternatively, fitting the dependence of the variance on DNA length to a linear dependence that includes a constant term yields fits that would imply large negative values of the variance at zero DNA length, which is unphysical. If we instead fit a linear dependence without an offset, we obtain a poor fit (Fig. 5.4b, dashed line; χ^2 = 13.6). If the complete label-label variance is attributed to the stretch modulus of the DNA alone, the slope of the linear fit implies a value for the stretch modulus of $S_{apparent}$ ~140 pN, considerably smaller than the value found in single-molecule stretching experiments.

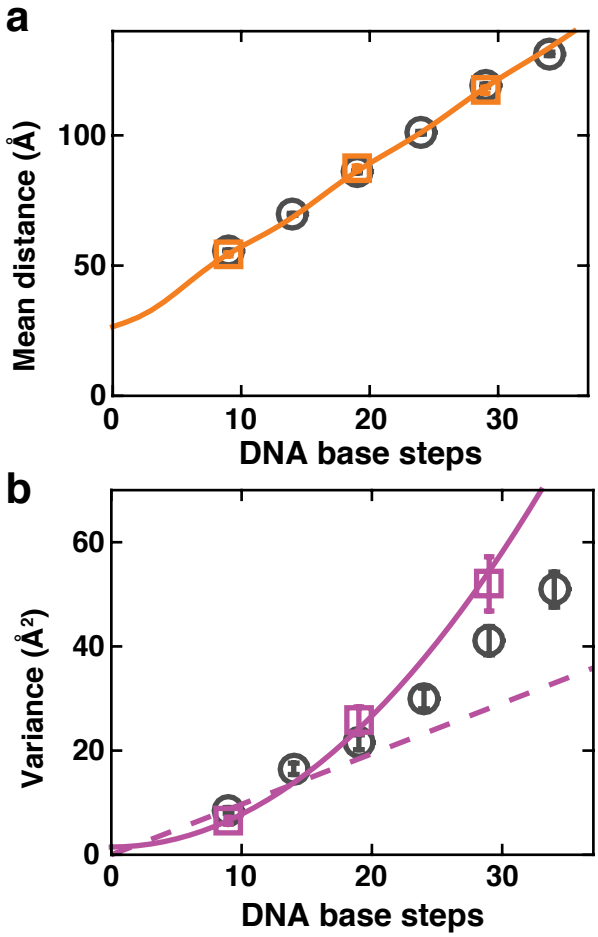

Figure 5.4: Distance parameters for double-stranded DNA obtained from ASAXS measurements. (**a**) Mean label-label distances obtained from the main peak of the distance distributions as a function of DNA base steps. Data obtained from ASAXS measurements described in this work (orange squares) and the best fit of a model (see Supporting Information 5.4) including the DNA length and label positions (orange line). Error bars are smaller than symbols. (**b**) Variances of the main peaks of the label-label distance distributions from ASAXS analysis (magenta squares). The data are well described by a quadratic dependence (χ^2 = 0.61; solid line) and incompatible with a linear dependence (χ^2 = 13.6; dashed line). For comparison the previously determined values [82] for the same constructs are shown as black circles.

The distance distributions obtained using the ASAXS approach (Fig. 5.3c, solid lines) are in excellent agreement with the results of prior measurements employing single-energy SAXS measurements and subtraction of single-labeled and unlabeled DNA contributions for the same DNA constructs and labels [82] (Fig. 5.3c, dashed lines and Fig. 5.4). Therefore, our results provide a clear confirmation, through an independent experimental approach, of the surprising findings obtained previously using SAXS at a single X-ray energy.

5.3 Conclusion

In summary, we have used ASAXS measurements at different energies around the gold L-III edge to determine the full label-label distance distribution for gold nanoparticles attached to the ends of 10-30 bp long DNA constructs. Our results demonstrate that by using small (~7 Å radius; ~80 gold atoms) nanoparticles, the label-label term can be reliably obtained from ASAXS measurements, which was not possible in previous measurements using single atom labels [48, 125, 144, 145]. Model calculations suggest that for molecules in the size range investigated here, even smaller gold labels than were used in this study, down to ~20 atoms, would be sufficient to obtain interpretable gold-gold interference patterns (Supplementary Fig. 5.6). Significantly larger molecules will require larger gold labels; to maintain an approximately constant level of relative anomalous scattering signal, the number of gold atoms in the labels should be increased proportionally to the number of atoms in the labeled molecule (Supplementary Fig. 5.6).

Our ASAXS method has the advantage that only the double-labeled sample needs to be prepared and measured and that it does not rely on intrinsic assumptions about the macromolecular scattering contributions being negligible; these properties will be particularly advantageous for macromolecular samples where selective labeling at only one position is difficult to achieve and/or that are strongly scattering. These advantages make our method reliable and experimentally attractive, yet it retains the full ability to provide absolute distances with Ångström resolution and precise distance distributions to evaluate the flexibility of macromolecular systems. While this proof-of-concept work uses gold-labeled DNA samples, we fully expect our method to be equally applicable to labeled proteins and to protein-nucleic acid complexes; similarly, other labels, such as silver and platinum nanoparticles, can provide equally attractive and orthogonal labeling options and might permit to measure several distinct distance distributions for one sample [139]. In conclusion, ASAXS provides a powerful new approach to determining intramolecular distance distributions for labeled biological macromolecules that we anticipate to provide new and quantitative insights into the structure, dynamics, and interactions of biological macromolecules.

5.4 Supplementary Methods

5.4.1 Sample preparation

Double-stranded DNA molecules 10, 20, and 30 bp in length and labeled at both ends with 7 Å radius thioglucose passivated gold nanoparticles were prepared as described previously [69, 82]. Briefly, water-soluble gold nanoparticles were synthesized as previously described [114]. Singlestranded oligonucleotides (ssDNA) were prepared on an automated ABI 394 DNA synthe-

sizer and thiols were incorporated using a C3-thiol-modifier (Glen Research, part #20-2933-41). ssDNA was purified by high-pressure liquid chromatography (HPLC) and coupled to gold nanoparticles in a 5:1 ratio (gold nanoparticle:ssDNA) at room temperature in 100 mM Tris-HCl, pH 9.0 for two hours. Gold nanoparticle-oligonucleotide conjugates were purified by ion-exchange HPLC and stored at −20 °C. Concentrations were determined by absorbance using a NanoDrop ND-1000 (NanoDrop Technologies). The DNA sequences used in this study are reported in Table 5.1. All SAXS measurements used 70 mM Tris-HCl buffer, pH 8.0, with 100 mM NaCl and 10 mM mM ascorbic acid added.

5.4.2 ASAXS Measurements and Data Analysis

X-ray energy calibration

ASAXS experiments were performed at the BESSRC-CAT beamline 12-ID of the Advanced Photon Source [140] (APS). Before starting an ASAXS measurement series, the experimental energy was calibrated by scanning the X-ray energy in steps of 1 eV around the L-III absorption edge of gold with a 50 µm thick gold foil inserted into the beam path and recording the incident I_0 and transmitted I_t X-ray intensity. The position of the absorption edge was determined from the maximum of the numerical derivative of $log(I_0/I_t)$. The absorption edge position was referenced to the tabulated `(http://henke.lbl.gov/optical_constants/)` value for the L-III absorption edge of 11.919 keV by applying an additive (and small, typically 5-10 eV) offset to all data recorded during that measurement run. All energies reported in this work have been calibrated by this offset, unless otherwise noted.

ASAXS measurements

All ASAXS measurements were carried out at room temperature using a ~16 µL sample cell and sample holder [19], a sample to detector distance of 1.0 m, a CCD detector (Mar CCD165), and sample concentrations of 200 µm. For each gold-labeled DNA sample, 10 measurements of 2 s exposure time were obtained at different energies in the range from 11.719 keV (200 eV below edge) to 11.969 keV (50 eV above edge). Control measurement using identical samples and exposure times at one energy exhibit no sign of radiation damage as repeat exposures are superimposable, within experimental error (Supplementary Fig. 5.7). Data were circularly averaged and buffer profiles recorded using identical procedures were subtracted for background correction. The scattering profiles recorded at different energies were interpolated to a common s-scale that comprised the lowest s-value of the lowest energy and the largest s-value of the highest energy with 1000 equally spaced increments in between, therefore avoiding extrapolation beyond the recorded s-value range for all energies.

Normalization of scattering profiles

We normalized the scattering profiles measured at different energies as follows. First, we obtained the values of $f_0 + f'(E)$ and of $f''(E)$ for gold from the Lawrence Berkeley X-ray data site `(http://henke.lbl.gov/optical_constants/sf/au.nff)`. Second we interpolate the values for $f_0 + f'(E)$ and $f''(E)$ to 1 eV resolution. For the imaginary part $f''(E)$ we employed simply linear interpolation. To achieve the same resolution for the

real part $f_0 + f'(E)$ of the gold atomic scattering factor, we fitted the following function to the tabulated values (Supplementary Fig. 5.7a):

$$f_0 + f'(E) = Re[a \cdot ln(E_{edge}^2 - E^2) + b \cdot ln(E_{edge}^2 + E^2)] \tag{5.5}$$

with E_{edge} = 11 918.6 eV. The fitting parameters a and $b$were determined to be 1.915 and 2.033, respectively. Third, we calculated theoretical energy-dependent intensity scaling factors $int(E)$ for each measured energy using the interpolated atomic scattering factors as

$$int(E) = |f_{Au}(E)|^2 + 2 \cdot Re(f_{Au}(E)) \tag{5.6}$$

Forth, we computed the raw mean scattering intensity for each of the experimental scattering profiles (Supplementary Fig. 5.7b) by averaging the scattering intensity over all s-values (Supplementary Fig. 5.7c, red circles). Fifth, the theoretical intensity factors $int(E)$ were multiplicatively scaled to match the raw intensities at the minimum intensity (which corresponds to the edge). Finally, all scattering profiles were normalized by multiplying by a factor equal to the ratio of the scaled theoretical intensity factor to the raw intensity (Supplementary Fig. 5.7c, green circles). Last, we applied an additive constant to the scattering profiles to match the normalized intensity profiles at high angles ($s \geq 0.08$ Å^{-1}) to correct for fluorescence [48].

Matrix inversion to obtain partial scattering factors

We determine the gold label-gold label interference pattern $G_{Au-Au}(s)$ by matrix inversion of Eq. 5.3 from the experimentally determined scattering profiles at the various energies (assembled in the matrix I) and from the matrix T that contains information about the energy and angle dependence of the scattering factors following the approach outlined by Pinfield and Scott [139]. The matrix equation is more explicitly visualized in Supplementary Fig. 5.8a.

The $(N_E \cdot N_s) \times (3 \cdot N_s)$ matrix T comprises the pre-computed $a_i(s)$, $b_i(s)$, and $c_i(s)$ that are all $N_s \times N_s$ square diagonal matrices containing label-label pair scattering factors, label-molecule pair scattering factors and intra-molecule pair scattering factors, respectively. The matrices $a_i(s)$ comprise the energy and angle dependent squared scattering factor of the gold label (Supplementary Fig. 5.8b). The individual matrices $a_i(s)$ for each energy E are constructed by multiplying the absolute value squared of the atomic scattering factor $|f_{Au}(E)|^2$ with a matrix that has the angle-dependent squared scattering factor of the gold label on the diagonal. To compute the scattering factor of the gold labels, we approximate the gold nanocrystal scattering by the scattering factor of a sphere with N_{Au} atoms and a radius R:

$$F_{Au}(s) = N_{Au} \cdot \frac{\sin x - x \cdot \cos}{x^3} \tag{5.7}$$

with $x = 2\pi \cdot R \cdot s$; for the gold nanocrystals used here [69] N_{Au} = 78 and R = 7 Å. In practice, we obtain the best fits when using a distribution of spheres, distributed uniformly with a spread of ±4% around R = 7 Å. The matrices $b_i(s)$ have the angle-dependent gold scattering factor $F_{Au}(s)$ on the diagonal and are multiplied by the real part of the gold atom scattering factor $Re(f_{Au}(E))$ for each energy (Supplementary Fig. 5.8c). Finally, the matrices $c_i(s)$ are constructed by multiplying the $N_s \times N_s$ identity matrix by the absolute value squared of the average atomic scattering factors of atoms in the DNA molecule at the respective energies, averaged over the atomic composition of our DNA constructs (Supplementary

Fig. 5.8d). The atomic scattering factors for all relevant atoms and energies were obtained from (http://henke.lbl.gov/optical_constants/). Before matrix inversion, the matrix T is conditioned by dividing all entries of the matrices a, b, and c by their respective elements corresponding to the smallest scattering angle and highest energy. The vector of partial scattering vectors G was computed by matrix inversion of Eq. 5.4 using the matrix left division in Matlab, which determines the least square solution of the system of linear equations.

Inversion of the label-label interference pattern to obtain label-label distance distributions

We obtain the distance distributions $P(d)$ by inverting the gold label-gold label structure factor $G_{Au-Au}(s)$ using the maximum entropy algorithm described previously [69, 82]. In brief, the gold label-gold label distance distribution $P(d)$ is determined at discrete values d by decomposing the label-label interference pattern into a linear combination of scattering interference profiles $I_d(s)$, which correspond to pairs of gold nanocrystals at a distance d:

$$G_{Au-Au}(s) = \sum_{d=0}^{d_{max}} P(d) \cdot I_d(s) \tag{5.8}$$

The sum runs over values of d from 1 to d_{max} in steps of 1 Å, where d_{max} was chosen to be approximately twice the mean label-label distance for each data set. The $I_d(s)$ are proportional to $\sin(2\pi \cdot d \cdot s)/(2\pi \cdot d \cdot s)$ with a prefactor that corresponds to the scattering of a pair of gold nanocrystals at zero distance and was taken from Mathew-Fenn et al. [69].

In principle, the values of $P(d)$ at each value of d can be determined from Eq. 5.8 using non-negative least squares. However, as was shown previously, a maximum entropy regularization procedure for the fit yields better results [69, 139] and was employed for the results presented in Fig. 5.3c.

Analysis of label-label distance distributions

The mean and variance of the gold label-gold label distributions was determined from analysis of the main peaks using the autopeak routine in Matlab. All data are provided in Table 5.1. Errors were estimated to be 1% for the length and 10% for the variance from triplicate repeat measurements.

The mean of the label-label distances was fitted to the model described previously (Supplementary Figure 2 of Ref. [82]) with the rise per base pair and the radial displacement of the labels as free fitting parameters. The axial offset $axial_0$ and angle offset θ_0 were kept fixed at their previously determined values of 24 Å and 1.34 π, respectively. The rise per base pair was determined from the fit to be of 3.23 ± 0.1 Å and the radial displacement to be 7 ± 1 Å.

5.5 Supplementary Tables

DNA construct	Sequence	Mean distance (Å) (This work)	Variance distance (Å^2) (This work)	Mean distance (Å) (Ref. [82])	Variance distance (Å^2) (Ref. [82])
10 bp	5'–GCATCTGGGC–3' CGTAGACCCG	54.3 ± 0.5	6.52 ± 0.7	55.7 ± 0.3	8.5 ± 0.6
20 bp	5'–CGACTCTACGGCATCTGCGC–3' GCTGAGATGCCGTAGACGCG	87.3 ± 0.9	25.8 ± 2.6	86.0 ± 0.4	21.6 ± 1.4
30 bp	5'–CGACTCTACGGAAGGTCTCGGACTACGCGC–3' GCTGAGATGCCTTCCAGAGCCTGATGCGCG	117.0 ± 1.2	52.0 ± 5.2	119.1 ± 0.6	41.1 ± 2.7

Table 5.1: DNA sequences and the mean and variances of the main peak in the gold label-label distance distributions obtained using the ASAXS method in this work and previously using single-energy SAXS.

5.6 Supplementary Figures

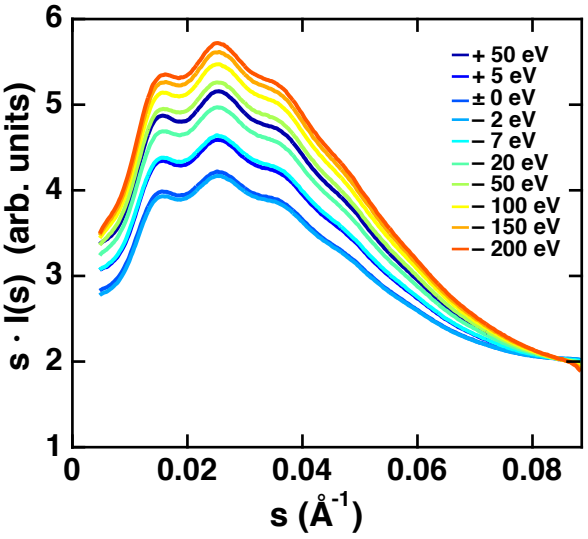

Figure 5.5: Scattering intensities as a function of momentum transfer s for the 20 bp double labeled DNA construct at 10 different energies shown in Holtzer representation of $s \cdot I(s)$ as a function of s that emphasizes differences in the intermediate s range (same data as in Fig. 5.3a). Energies in the figure legend are relative to the gold L-III edge.

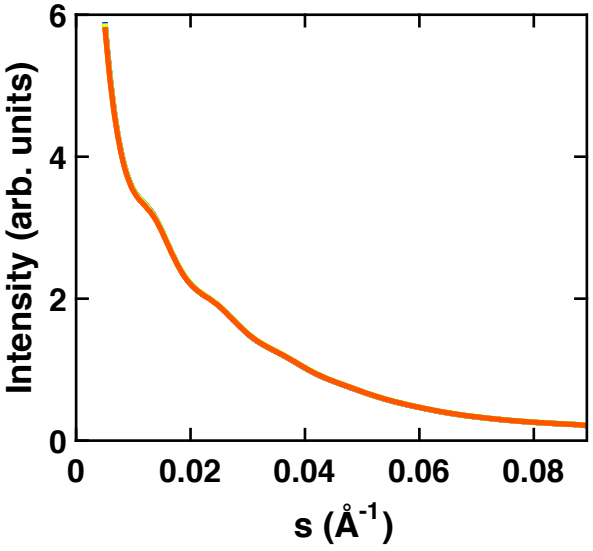

Figure 5.6: Scattering intensity as a function of momentum transfer s for the 20 bp double gold labeled DNA construct. 10 subsequent exposures (color coded from blue to red) of 2 s each all at the same X-ray energy of 11.911 keV (8 eV below the gold L-III edge) are shown. Scattering profiles are identical within experimental error, indicating the absence of radiation damage.

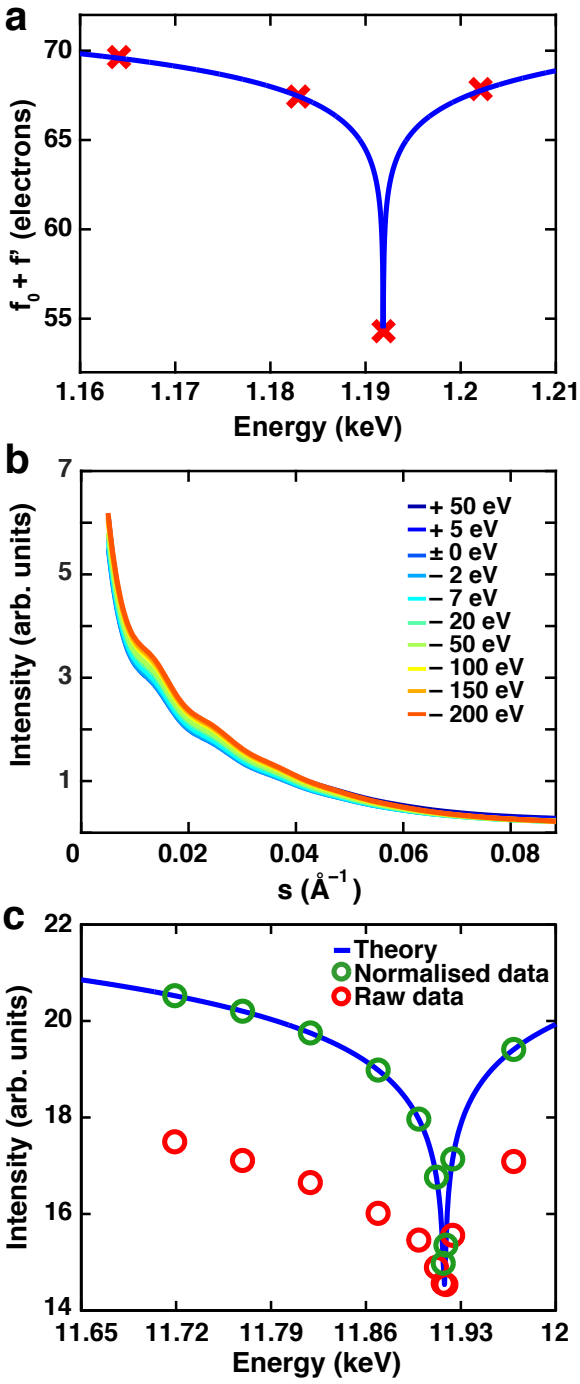

Figure 5.7: Normalization of ASAXS profiles. (**a**) Real part of the atomic scattering factors for gold close to the L-III edge. Crosses are the tabulated values obtained from `http://henke.lbl.gov/optical_constants/sf/au.nff`. The solid blue line as a fit of Eq. 5.5 to the data (see Methods for details 5.4). (**b**) Scattering intensities as a function of momentum transfer s for a 20 bp double labeled DNA sample at 10 different energies prior to normalization. The same data set after the application of the energy dependent normalization is shown in Fig. 5.3. (**c**) Normalization factors for the energy dependent normalization. Raw intensity values, obtained by averaging scattering intensities at each energy over all s values are show as red circles. The theoretical scattering intensity scaling curve (Eq. 5.6) adjusted to the minimum value in the raw intensity data is shown as a blue line. Scaled mean intensity values are shown as blue circles. For details of the normalization procedure see Methods 5.4.

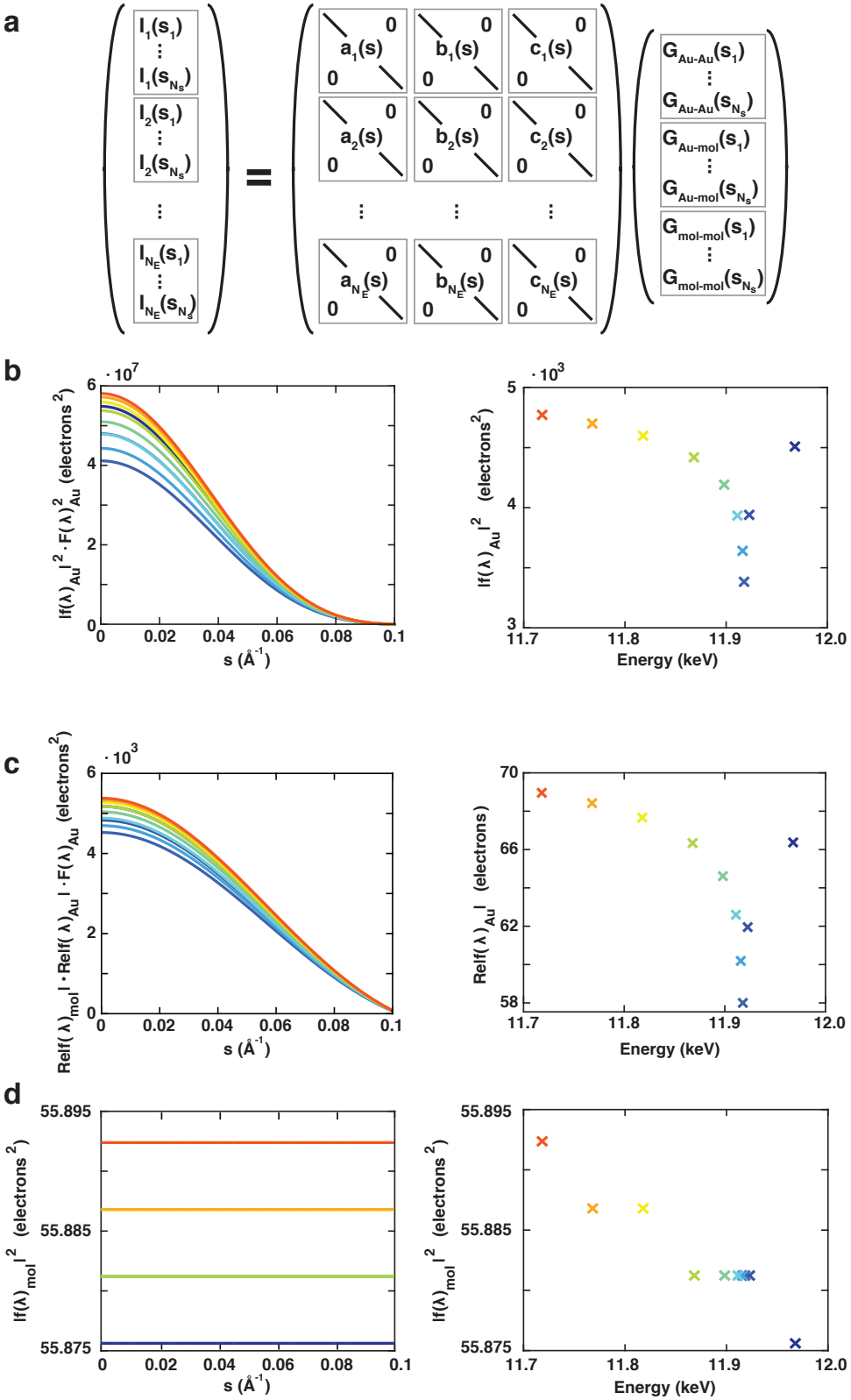

Figure 5.8: Explicit form of the matrices $a_i(s)$, $b_i(s)$, and $c_i(s)$ used in the matrix inversion. (**a**) Structure of the matrix equation (Eq. 5.4 in the main text) connecting the experimental measured scattering intensities $I_i(s)$ to the partial structure factors G_{Au-Au}, G_{Au-mol}, and $G_{mol-mol}$ by multiplication with the pair scattering factors of the

gold-gold terms a_i, pair scattering factors of the cross terms b_i, and pair scattering factors of the molecule-molecule terms c_i at energies E_i. (**b**) Plots of the diagonal entries of the a_i matrices, i.e. the pair scattering factors of the gold-gold terms a_i as a function of s (left), which are the scattering functions for the spherical gold labels with a radius of 0.7 nm (Eq. 5.7, squared) scaled by the absolute value squared of the atomic scattering factors of gold as function of energy (plotted on the right). (**c**) Plots of the diagonal entries of the b_i matrices, i.e. the pair scattering factors of the gold-molecule cross terms (left). The s dependency is caused by the spherical gold label scattering factor (Eq. 5.7); the energy dependence is given by real part of the gold atomic scattering factors (right). (**d**) Plots of the diagonal entries of the pair scattering factors for the molecule-molecule terms c_i (left). There is no dependence on s, therefore, the scattering factors are horizontal lines (left). The energy dependence is given by the average of the atomic scattering factors for DNA molecules as function of energy (right). Since the atomic scattering factors for the elements making up biological samples are almost constant in the energy range considered, the c_i show almost no energy dependence (compare the y-scales of the right plots in panels (b), (c), and (d)).

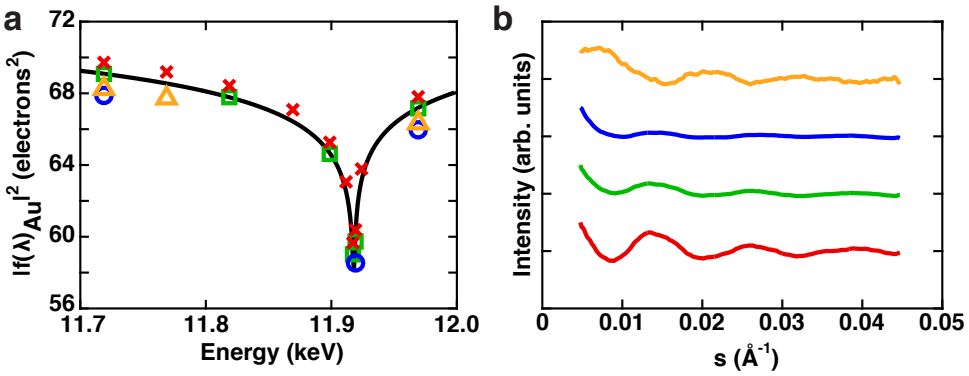

Figure 5.9: Stability of the partial structure factor computation using scattering profiles at different energies. Using the data set shown in Fig. 5.3 of the main text as a starting point, we selected sub-sets of the scattering profiles at different energies and computed the gold-gold interference term by inversion of Eq. 5.4 as described in the main text. (**a**) Atomic scattering factor of gold as a function of energy (black line). The sets of energies selected for different calculations of the gold-gold interference term are shown as differently colored symbols. The symbols are slightly vertically offset for clarity. (**b**) Gold-gold partial structure factors computed using sub-sets of the scattering profiles at different energies. The color code corresponds to the symbols in panel (a). Profiles are vertically offset for clarity. Using 10 different energies (red crosses in panel a), i.e. the full data set, for inversion yields a well defined partial structure factor (red line in panel (b)), as discussed in the main text. Using only 6 energies (green squares in panel (a)) or 3 energies (blue circles in (a)) positioned below, on, and above the absorption edge still enables inversion of the matrix equation and gives rise to gold-gold partial structure factors with the expected oscillating pattern but with (much) reduced amplitudes (green and blue lines in panel (b)). Using scattering profiles at 3 energies away from the gold L-III absorption edge (dark yellow triangles in panel (a)) -and therefore with little change in scattering intensity as a function of energy- is insufficient for robust inversion of the matrix equation; the resulting gold-gold partial structure factor is very noisy and shows spurious oscillations (dark yellow line in panel (b)).

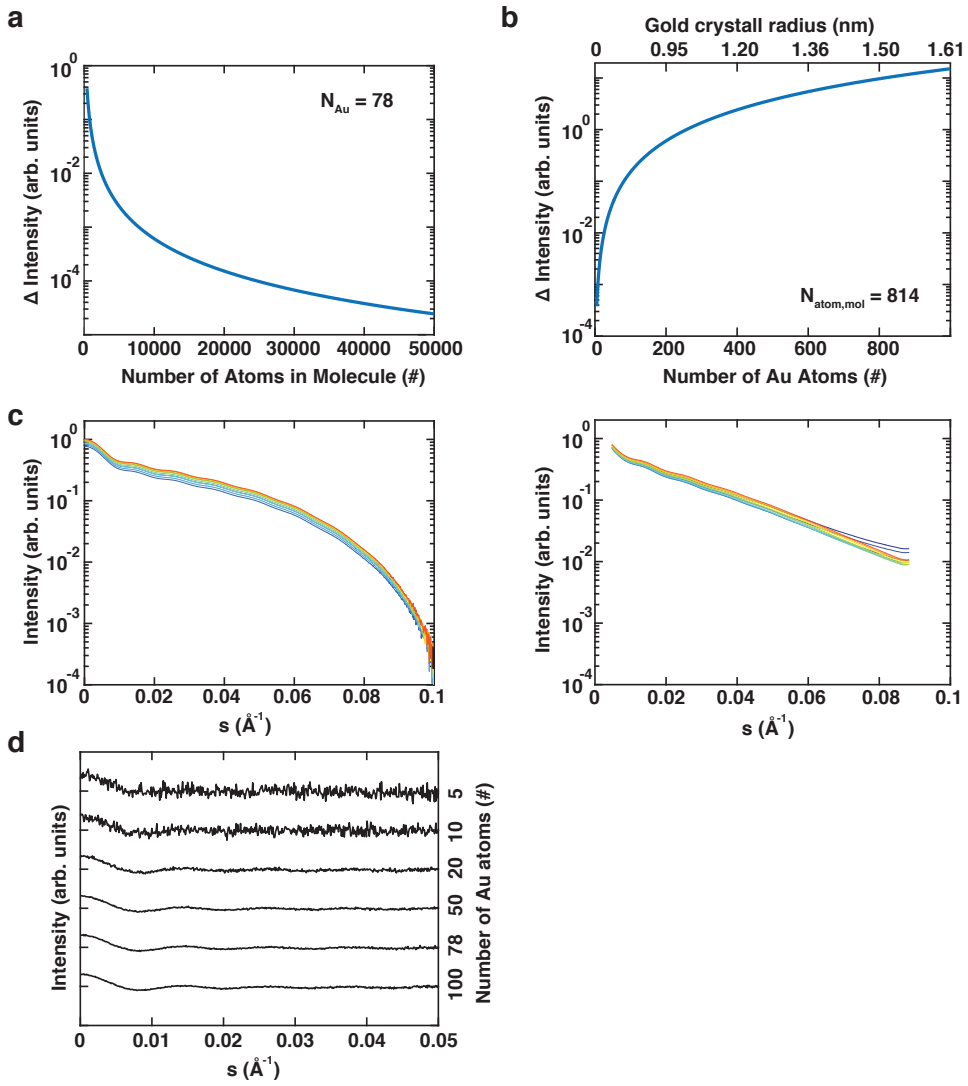

Figure 5.10: Relative scattering intensities of the gold labels and molecules. The level of the anomalous signal from gold nanocrystals relative to the scattering from the molecule is evaluated by the relationship

$$\Delta I = \frac{N_{Au}^2 \Delta |f_{Au}(\lambda)|^2}{N_{mol}^2 < |f_{mol}|^2 >} \tag{5.10}$$

where N_{Au} is the number of gold atoms in the nanocrystals, N_{mol} is the number of atoms in the labeled macromolecule, $< |f_{mol}|^2 >$ is the absolute value squared of the average atomic scattering factor of the macromolecule (averaged over the element composition of the molecule), and $\Delta |f_{Au}|^2$ is the change in the absolute value squared of the atomic scattering factor for gold, evaluated on and off edge.
(**a**) Ratio between the energy dependent variations of the gold scattering term and the molecular scattering contribution (computed using Eq. 5.10) as a function of the size of the macromolecule (in atoms). The number of gold atoms in the label was kept constant at 78, corresponding to the size of labels used in this study. (**b**) Ratio between the energy dependent variations of the gold scattering term and the molecular scattering contribution (computed using Eq. 5.10) as function of the number of gold atoms in the label. The corresponding approximate gold nanocrystal radius is indicated in the top axis. The number of atoms in the molecule was kept constant at 814 atoms, corresponding to the 20 bp DNA used in this work. Based on our experiments, we can estimate that a level of $\Delta I \sim 0.1$ is sufficient for reliable inversion; lower values of the intensity variation ΔI might still be sufficient for the computation of distance distributions for high quality measurements.

(**c**) Simulated intensity profiles for double labeled 20 bp DNA molecules, assuming gold labels with 78 gold atoms, with random noise of 0.01% I_0 added to the scattering profile (left) and experimentally measured profiles for the double labeled 20 bp DNA (right). The simulations used DNA structures generated with the make-na server (`http://structure.usc.edu/make-na/server.html`) and followed the approach of Pinfield and Scott [139]. The energy color code is the same as in Fig. 5.3. While random errors with a fixed magnitude of 0.01% I_0 do not completely describe the error encountered in the real experiments, this level of error is approximately representative of our experimental noise. (**d**) Simulated label-label partial structure factors for double-labeled 20 bp DNA with different gold label sizes as a function of s. Simulations were carried out as shown in panel (c), for different label sizes; label-label partial structure factors were computed with the same routines as used in the analysis of the experimental data. The amplitude of the oscillating interference pattern is decreasing with decreasing label size (bottom to top), while the noise level is visibly increasing.

Chapter 6

Remeasuring the Holliday Junction

Summary

The DNA four-way (Holliday) junction is the central intermediate of genetic recombination, but despited being extensively studied several key aspects of this important motif remain unclear. Although transition dynamics between alternative stacking conformations have been demonstrated, conventional single molecule studies cannot resolve the fast transitions at physiological salt. Moreover, while comparative gel electrophoresis has been used to test structural models under low or intermediate salt the precise solution conformational ensemble and configuration still remains unclear. Using a emerging technique termed X-ray scattering interferometry (XSI) we have been able to reveal structural and ensemble properties of the Holliday junction under various salt concentrations. Our results demonstrate that the four-way junction does not adopt a fully planar conformation under low ionic strength. Further, data recorded at intermediate salt conditions suggest a novel folding state. This information will be necessary to improve our understanding of DNA four-way junctions and evaluate existing models.

6.1 Introduction

The importance of Holliday junctions as fundamental nucleic acid structure motives that play central roles in genetic recombination and other cellular processes. Additionally they developed into a widely used tool in DNA nanotechnology [146–151]. Holliday junctions, and nucleic acid junctions in general, rely on their abilities to undergo conformational changes to carry out their biological functions either alone or as building blocks of larger nucleic acid and nucleic acid-protein complexes [150, 152–156]. There are many proteins such as junction-resolving enzymes that can recognize and distort the structure of the junction, for example by stabilizing the unstacked conformation, breaking the central base pairs or changing the interduplex angle between the arms of the stacked conformation [150, 153, 154]. However, these enzymes are highly specific for the structure of the junction [156]. Thus, defining the underlying conformational landscapes or ensembles of Holliday junctions and other DNA/RNA junctions are required to fully understand the action of proteins that modify the conformational landscape.

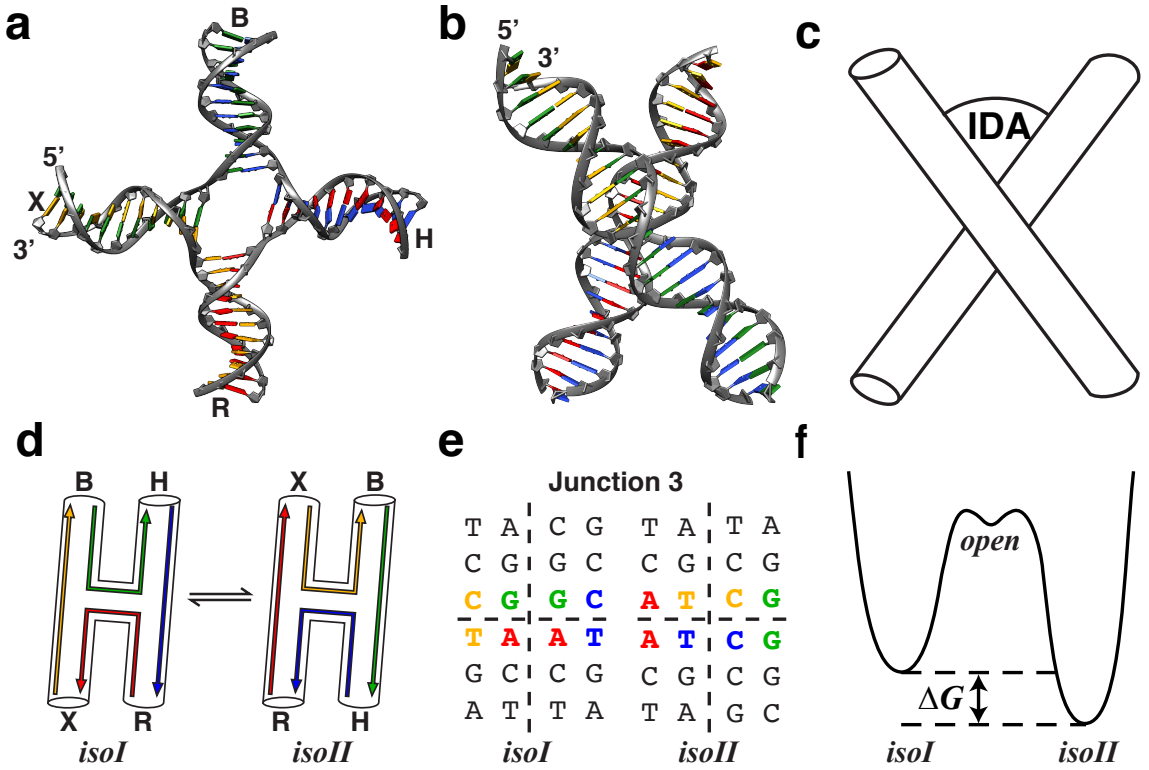

Figure 6.1: Structure of the Holliday junction. (**a**) Schematic of a four-way junction in the extended open conformation observed in the absence of added metal counter-ions with the helical arms X, B, H and R. Crystal structure obtained from [164] (PDB Code: 2CRX). (**b**) Schematic of the stacked X-structure observed in the presence of counter-ions. (**c**) Two right-handed helices are folded by coaxial stacking with an inter-duplex angle (IDA) on the order of ~60° between arms [150, 161, 165]. (**d**) Schematic of the two anti-parallel conformations *isoI* and *isoII* of the X-structure with different orientation of the helical arms X, B, H and R. (**e**) The central sequence of the chosen junction motif 3 in both stacked states. The different colors visualize the individual pathways of the oligonucleotides between the two states. (**f**) The proposed energy landscape of conformer transitions with two distinct stacked states and an intermediate open state [155, 160].

Experimentally, it is known that the open state is predominant at low salt concentrations, and, at higher ionic strength, junctions adopt stacked conformations (Fig. 6.1a-c) [152, 155, 157–160]. The stacked conformations, which are induced through the binding of metal ions, lead to the formation of two quasi-continuous helices with pairwise stacking of the helical arms (Fig. 6.1b-e). There are two ways of forming a stacked structure differing in choice of stacking partner given by the central motif of the junction (Fig. 6.1d-f) [155, 157, 161]. The choice of stacked conformation seems to play a key role in genetic recombination as there is evidence that this can influence the binding orientation of junction-resolving enzymes and therefor alter the cleavage positions [154, 156, 157, 162, 163].

While the conformational landscapes of Holliday junctions have been extensively studied, many open questions remain. Single-molecule förster resonance energy transfer (smFRET) has identified multiple states under solution conditions with high Mg^{2+} concentrations where the Holliday junction's conformational changes are slow enough for conventional smFRET detection (Fig. 6.1f) [155, 159, 160]. Although förster resonance energy transfer (FRET) values of these Holliday junction states do not provide direct information on the distances or on whether each of them are single rigid states or a conformational ensembles, single molecule recordings

can be used to obtain useful information about reaction rates of individual molecules. We incorporated dyes at targeted ends of the Holliday junction to estimate the relative populations for the classical two-state model and provide information about transition dynamics using smFRET. Moreover, we applied an emerging synchrotron based structural technique, X-ray Scattering Interferometry (XSI), to resolve the conformational ensemble of the Holliday junction and to obtain precise structural and dynamical information. To answer specific questions including *i*) how the distribution of Holliday junction states change with salt; *ii*) to separate and identify different molecular conformations; and *iii*) what are the precise solution structure and dynamical properties of the major Holliday junction conformational states we studied a model Holliday junction sequence under different salt conditions. We used these experimental data to generate and test physical models of Holliday junctions to demonstrate the change of the energy landscape under different counter-ion environments.

We chose to examine the well-studied Holliday junction type 3. Studies of this macromolecule allowed us to match our XSI data against prior published information obtained by a wide range of techniques from comparative gel electrophoresis [157, 161], atomic force microscopy imaging [166], smFRET [155, 159, 160], small-angle X-ray scattering [167] to theoretical studies using molecular dynamics simulations [168] and to demonstrate the capability of XSI to obtain unique whole ensemble distance distributions with Ångström resolution. Moreover, we can compare our results for the junction type 3 to junctions with previously published findings for different central motifs [150, 157, 158, 160, 169].

The conformational free energy landscape of Holliday junctions is highly textsalt-dependent [155, 159]. At low salt (e.g. 30 mM Tris-HCl buffer only) the electrostatic repulsion between the DNA phosphate groups forces the Holliday junction motif (such as Junction sequence 3) to remain in an open state, which is suggested to have the arms pointing towards the corners of a square [150, 157, 161] (see Fig. 6.1a). In contrast, at high salt (10 mM Mg^{2+}), Holliday junctions tend to stably adopt stacked conformations known as *isoI* and *isoII* [152, 155, 157, 159, 161] (Fig. 6.1b-d). At intermediate salt concentrations (~150 mM Na^{+} or ~100 µM Mg^{2+}), the electrostatic repulsion of the DNA backbone is still sufficiently screened to allow the junction to adopt stacked conformations [159, 161] (Fig. 6.1b-d).

6.2 Results

6.2.1 Single-Molecule FRET Results

We measured conformational changes of individual molecules in real-time to study the relative population and furthermore the transition between the two stacked states. We assembled the DNA junction from four synthetic oligonucleotides with a length of 22 bases, each (see Materials and Methods 6.4 for details). The oligonucleotides self-assemble into the desired four-way junction motif with four helical arms (X, B, H and R, Fig. 6.1a,b) which cannot undergo branch migration [155]. Two arms were conjugated with dyes at the 3' termini to form a FRET pair. By incorporating a Cy3b as donor (arm H) and a Cy5 as acceptor (arm R) to provide FRET readout. Furthermore, we modified one arm with a biotin (arm X) for surface immobilization (Fig. 6.2a). The junction has a large inter-dye distance in first stacked conformation *isoI* resulting in a low FRET value whereas it has a small the distance in the second stacked conformation *isoII* with a high FRET signal (Fig. 6.2a). Our data at 10 mM $MgCl_2$ supplemented with 30 mM Tris show

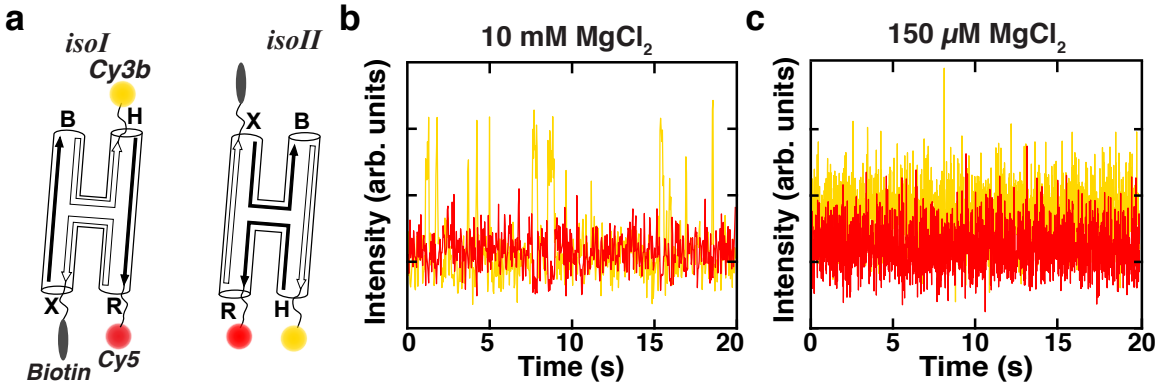

Figure 6.2: Single-molecule FRET layout and time recordings. (**a**) Schematic of the stacked states *isoI* and *isoII* of the Holliday junction including the incorporated dyes at the 3' termini of H (Cy3b) and R (Cy5) and a biotin label (arm X). (**b**) Typical smFRET trace recorded using a salt concentration of 10 mM $MgCl_2$ and 30 mM Tris and integration time of 3 ms. Transitions between the two states are nicely visible by the change in donor and acceptor intensity. (**c**) Typical smFRET trace recorded using following buffer 150 µM $MgCl_2$ and 30 mM Tris and integration time of 3 ms. No clear transitions are visible for the low salt condition.

a high FRET signal for a majority of the trace only interrupted by short low FRET intervals (Fig. 6.2b). The transition between the two states occurred within one frame (3 ms integration time, see Materials and Methods 6.4) and the signal show a heavy bias towards state *isoII*. The rates for the transition between the conformations $k_{I\to II}$ and $k_{II\to I}$ were averaged over hundreds of transitions from 77 individual traces and calculated to be 45.6 s^{-1} and 7.9 s^{-1} respectively and thus a population of 14.8 % for *isoI* and 85.2 % for *isoII* (Supporting Figs. 6.6 and 6.7). These results agree well with previously published population ratio of 1:4 from smFRET studies of junction with the chosen sequences [155, 159]. The strong preference for stacked state *isoII* can be explained by the difference in free energy contribution from the central base pairs (Fig. 6.1d). Moreover, the faster transition rates measured are in line with the dependence on the magnesium ion concentration and slower rates reported for 50 mM Mg^{2+} [155, 159].

In addition to the traces taken at 10 mM $MgCl_2$ we recorded data at lower ionic strength using a buffer containing 150 µM $MgCl_2$ and 30 mM Tris, pH 7.4. There are no clear hopping events between the donor and acceptor signals (Fig. 6.2c). This agrees with the model that the barrier between the two stacked states is lowered and thus the transitions between the two conformation occur on a much faster time scale as reported previously [155, 159]. Single-molecule FRET has the drawback that photons of a single dye must be integrated over a certain time to provide sufficient signal-to-noise. This limitation of standard smFRET prevents to resolve multiple populations in the free energy landscape of the Holliday junction under physiological buffer conditions since the transitions between the two mainly populated stacked states *isoII* and *isoI* are too fast to be resolved [155, 159].

6.2.2 X-ray scattering interferometry data results

In contrast to single-molecule FRET, X-ray scattering interferometry relies on ensemble averaging, but enables to extract high-resolution distances on an absolute scale and therefore structural and mechanical information about individual conformations even for very fast transition times. These valuable facts allowed us to study the Holliday junction at low salt

concentrations even in the transition regime where the electrostatic repulsion of the backbone charges get only partially screened. XSI measures the interference scattering pattern produced between two site-specifically attached electron rich markers together with scattering terms caused by the molecule and the cross-scattering terms between the labels and the macromolecule. This collective interference between the two labels is extracted from the remaining components and directly converted into an absolute distance distribution using regularized Fourier transformation [69, 82, 83, 105–108].

Similar to samples used in smFRET experiments we assembled the DNA junction from four synthetic oligonucleotides with a length of 22 bases each (see Materials and Methods for details 6.4). To provide the characteristic signal in XSI we attached various pairs of gold nanocrystal markers (HB, HX, XB, and RH) following the approach published previously [69, 82, 83, 106] (see Materials and Methods 6.4). Furthermore, we assembled, purified and recorded the corresponding single labeled and unlabeled constructs required to extract the gold-gold interference pattern. Two distinct conformational populations are expected for the designed gold nanocrystal pairs.

Fig. 6.3b shows the gold-gold distance probability distribution for the Holliday junction at high divalent salt (10 mM $MgCl_2$ and 30 mM Tris-HCl, pH 7.4) across three different label pairs (Fig. 6.3a). For convenience, we added schematics of the investigated gold label pairs in both the *isoI* and *isoII* (Fig. 6.3a). The varying attachment positions facilitated examination of orientations adopted by the junction arms to each other and allow for comparing the conformations in the different states.

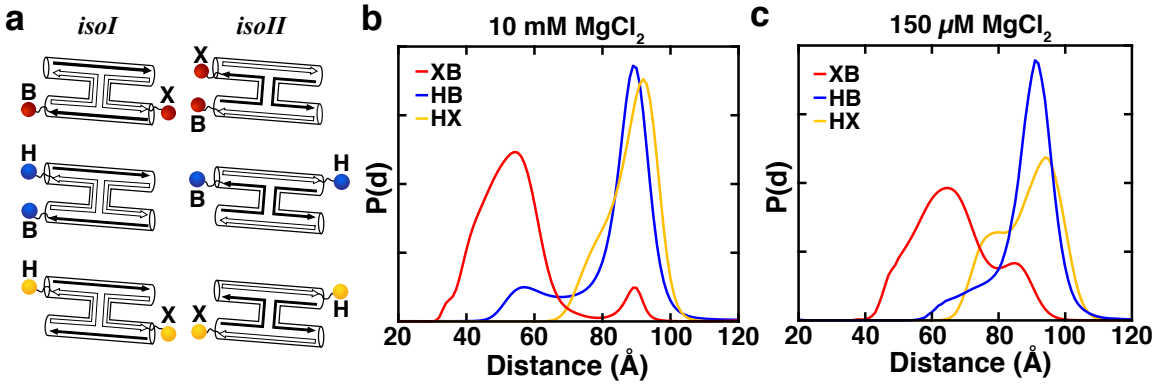

Figure 6.3: XSI layout recorded distributions at high and intermediate divalent salt. (**a**) Schematic of the stacked states *isoI* and *isoII* of the Holliday junction attached label pairs (XB, HB and HX). Distance distributions for the label pairs XB (red), HB (blue) and HX (yellow) recorded at (**b**) high (10 mM $MgCl_2$ and 30 mM Tris, pH 7.4) and (**c**) intermediate divalent salt concentration (150 µM $MgCl_2$ and 30 mM Tris, pH 7.4). The different states are nicely visible for both salt conditions. Traces in (b) and (c) were normalized to sum to unity.

The distance distribution is dependent on the ionic condition. Under conditions with high charge screening of 10 mM $MgCl_2$, we observe two peaks for all label pairs. While the peaks are clearly separated for the XB (Fig. 6.3b, red) and HB (Fig. 6.3b, blue) label pairs the distribution for HX (Fig. 6.3b, yellow) is more narrow. The major population for the XB label pair centers around 54 Å and a minor second population is located at 89 Å (Fig. 6.3b, red), whereas for the HB labeled version the main peak is at 89 Å and a second smaller peak at 57 Å (Fig. 6.3b, blue). These findings are in line with the proposed X-shape switching between two stacked conformers and FRET studies reporting a bias for *isoII* [155, 159]. It is striking that both XB

and HB show the exact same distance for the configuration were both labels are positioned at opposite ends of a continuous helix. However, the high resolution of XSI enabled us to detect a small difference between the short distance distribution of these label pairs. One possible consequence of this finding is that the junction adopts slightly different interduplex angles depending on the choice of stacking partner. Another important result at this conditions is the end-to-end distance of 89 Å for opposite labels. One would expect a displacement of ~93 Å assuming a continuous B-form DNA helix [150] with helical rise of 3.3 Å [82, 105] and gold label offset parameters reported for dsDNA previously [82, 105]. This suggests that the DNA helix is overwound or bent to some degree. Nevertheless, bending of the structure is more likely as there is significant electrostatic repulsion between the phosphates located along the backbone and theoretical studies proposed slight underwinding for bases close to the cross-over [169].

Moreover, we used the same gold nanocrystal pairs under 150 µM $MgCl_2$ and 30 mM Tris-HCl, pH 7.4 which allowed direct testing of how the Holliday junction motif 3 responds to intermediate divalent salt (Fig. 6.3c). In agreement with high divalent salt data the junction still prefers to adopt *isoII*. Moreover, the slight offset of the high distance peak (Fig. 6.3c, blue line) for label pair HB (*isoII*) as compared to high salt (Fig. 6.3b, blue line) agrees well with the reported small salt-dependent alteration in the gold nanocrystal position [107]. The shift of *isoII* of label pair XB (Fig. 6.3c, red line) is in line with lower electrostatic screening at intermediate salt conditions and therefore an increase in interduplex angle for a stacked conformer. However, the peak gold-gold distance for XB and HX in conformer *isoI* change notably. The displacement to lower distances for both peaks suggest that the junction does no longer arrange in a stacked structure but a different conformation in *isoI*. Potentially the gain in free energy for the stacking partners in *isoI* is not sufficient to overcome the electrostatic repulsion under this salt concentration. Our XSI results with a stacked conformer for the major population above ~100 µM Mg^{2+} agree well with prior gel electrophoresis [161] and bulk FRET studies [159]. Nevertheless, this is the first time that the full ensemble properties at intermediate divalent salt concentration are revealed showing that possibly only one stacked conformer is stable under these conditions.

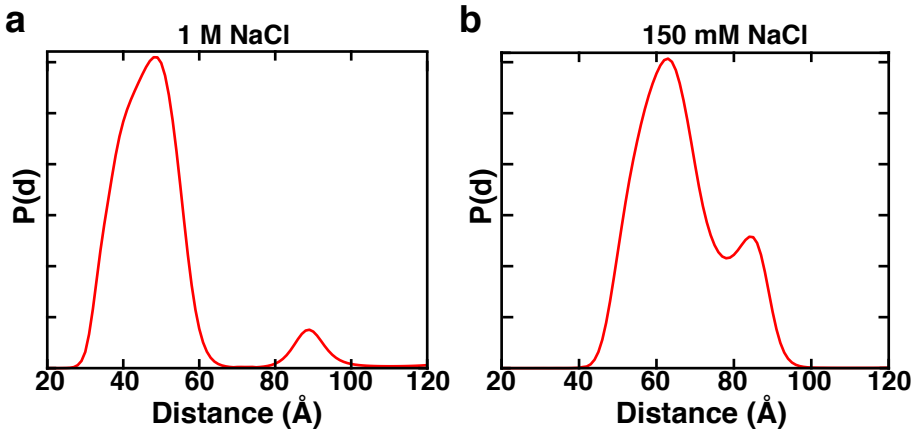

Figure 6.4: XSI layout recorded distributions at high and intermediate monovalent salt. Distance distributions for the label pair XB recorded at (**a**) high (1 M NaCl and 30 mM Tris-HCl, pH 7.4) and (**b**) intermediate monovalent salt concentration (150 mM NaCl and 30 mM Tris-HCl, pH 7.4). The different states are nicely visible for both salt conditions.

To test whether the Holliday junction responds specifically to Mg^{2+} we further recorded XSI data for high (1 M NaCl and 30 mM Tris-HCl, pH 7.4) and intermediate (150 mM NaCl and 30 mM Tris-HCl, pH 7.4) monovalent salt for the label pair XB similar to experiments with $MgCl_2$ (Fig. 6.4). The results for sodium show a similar behavior as the profiles obtained with magnesium. We therefore conclude in agreement with analysis of conformational species by FRET [159] that the junction adopts the same two-state behavior (*isoI* and *isoII*) as for divalent salt with two well stacked states at high salt (Fig. 6.4a), although much higher concentrations are required. Moreover, data recorded at 150 mM NaCl indicate only one fully stacked state (Fig. 6.4b) equally to profiles at 150 µM $MgCl_2$.

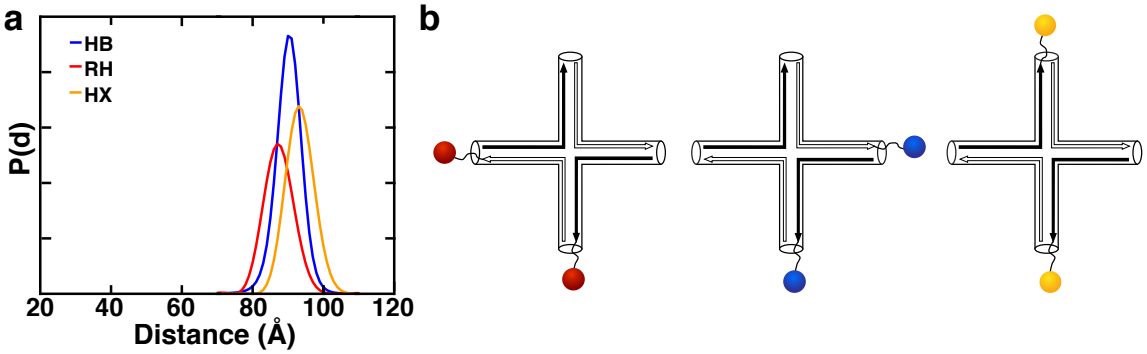

Figure 6.5: XSI layout recorded distributions at low salt. (**a**) Distance distributions for the label pairs HB, RH and HX recorded at low salt (30 mM Tris-HCl, pH 7.4). Distance distributions were normalized to a total probability of 1. (**b**) Schematic of the proposed open state at low salt with the label pairs from (a).

To analyze the salt dependency further and to access weather the unstacked conformation at intermediate salt corresponds to the proposed square open configuration we probed gold-gold distances under low salt conditions (30 mM Tris-HCl, pH 7.4). As expected, our XSI experiments showed that the junction cannot from the stacked X-structure as there is only minor charge screening without Na^+ or Mg^{2+} ions present (Fig. 6.5). These findings are consistent with prior FRET and gel-mobility data [150, 161, 170]. The peak gold-gold distance varies slightly for the different label pairs. Moreover, the results provide evidence for a pyramidal structure as the measured distance for HX (94 Å, Fig. 6.5) is too small for a fully planar conformation with an estimated displacement of ~111 Å. The theoretical value was calculated using a gap of 16.3 Å for the center predicted from a protein-bound planar crystal structure [164], a helical rise of 3.3 Å and gold nanocrystal axial and rotational offset parameters from Mathew-Fenn *et al.* [82]. While this only provides a rough approximation as the junction center is slightly distorted by the bound protein the experimentally determined distance deviates too much to support the model of a plane structure. Given the two faces of DNA, namely major and minor groove, the arrangement of these inequivalent sides might lead to a pyramidal conformation. Another finding of the XSI data are the different peak positions for RH and HB (Fig. 6.5). This suggest a not fully symmetric alignment of the arms as this would result in identical distances. Comparative gel-mobility studies from Duckett *et al.* [161] support this finding showing asymmetric migration for pairwise extended arms. Again, the interaction of major and minor groove faces can be accountable for this effect as well as the individual base-pairs in the center.

6.3 Discussion

We have used XSI measurements to investigate the solution ensemble of the Holliday junction motif 3. Our XSI results agree well with prior FRET and gel-mobility studies for high salt, showing that the four-way junction preferentially adopts two distinct X-structures with coaxial pairwise stacking of helical arms. Further, XSI has allowed us to uncover conformational distributions at intermediate and low salt conditions. The results underscore the ability of XSI to reveal precise ensemble properties across a wide range of ionic strength even for rapidly interchanging conformations. Moreover, the Holliday junction is by now the largest macromolecule studied using XSI underscoring the scalability and potential of this method.

Although XSI lacks the all-atom structural precision attainable by other techniques, it can provide accurate solution and ensemble information. As XSI provides highly precise distance distribution information, it may be used to refine molecular dynamics simulations to gain full structural details.

Our results suggest that various solution condition as well as the bases located in the center of the structure can dramatically alter the free-energy landscape of the Holliday junction as stable formation of only one stacked state was observed at intermediate salt. Thus, generalization to other junction motifs is not possible straightaway since the motifs differ in central sequence. However, our XSI experiments showed that the sequence promotes distinct conformations under intermediate salt conditions. One possibility is that the binding orientation of junction-resolving enzymes and therefor genetic recombination can be regulated that way.

While this study used a single central motif, we expect our method to be readily applicable to other sequences to probe the free-energy landscape for multiple four-way junctions. Moreover, investing macromolecules with junction-specific proteins bound can provide testable models for the underlying behavior required for genetic recombination.

6.4 Materials and Methods

6.4.1 Preparation of Gold-DNA Conjugates

Gold nanocrystals with particle radius of 7 Å were synthesized and purified as described previously [69, 82, 83]. While unmodified single-stranded DNA (ssDNA) was purchased from Integrated Device Technology (IDT), C3-thiol-modified oligonucleotides were synthesized at the Protein and Nucleic Acids Facility at Stanford University on an automated ABI 394 DNA synthesizer. All DNA sequences used in this study are reported in Tables 6.1 and 6.2. Furthermore, ssDNA were purified by ion exchange high-pressure liquid chromatography (HPLC) using a Dionex DNAPac 200 column and a salt gradient from low (10 mM NaCl) to high salt (1.5 M NaCl) supplemented with 20 mM sodium borate, pH 7.8. Gold-ssDNA conjugates were formed by mixing oligonucleotides with gold nanocrystals in a 1:5 ratio (ssDNA : gold nanocrystals) at room temperature in 100 mM Tris-HCl, pH 9.0 for two hours. Immediately after the reaction conjugates were purified by HPLC using the Dionex anion exchange column and an elution gradient from 10 mM to 1.5 M NaCl supplemented with 20 mM ammonium acetate, pH 5.6 and desalted using Amicon centrifugal filters with 3 kDa cutoff (3500 × g with a swinging bucket rotor for 35 min; three repeats) at 4 °C. DNA dimers (e.g. R and X form RX) were hybridized for 30 min at room temperature follow by HPLC

purification using the same protocol as for the ssDNA gold nanocrystal conjugates. The full Holliday junction with two gold labels, one gold label and the unlabeled version were formed by mixing two equimolar complementary dimers (e.g. RX mixed with HB) in 10 mM $MgCl_2$, 10 mM Tris-HCl, 1 mM EDTA, pH 8.0 and annealed by heating the solution to 65 °C for 2 min followed by 39 °C for another 10 min in a thermocycler (BioRad). Lastly, the fully assembled structures were purified similar to DNA dimers, desalted using 10K Dalton Amicon filters (3500 × g with a swinging bucket rotor for 15 min; three repeats) at 4 °C and stored at −20 °C until SAXS experiments. After each desalting step concentrations were determined using a NanoDrop ND-1000 (NanoDrop Technologies) measuring the UV absorbance at 260 nm and 360 nm.

6.4.2 SAXS Measurements and Data Processing

Small-angle X-ray scattering experiments were carried out at beamline 4-2 of the Stanford Synchrotron Radiation Ligthsource and BM29 at the European Synchrotron Research Facility in Grenoble. At beamline 4-2 data was collected using a detector distance of 1.1 m at 11 keV whereas at BM29 the detector distance was set to 2.867 m with a X-ray energy of 15 keV. At both beamlines the experimental conditions were set to 30 µM sample concentration. Moreover, recordings were performed at 15 °C at beamline 4-2 and 5 °C at beamline BM29. A basic buffer containing 30 mM Tris-HCl, pH 7.4 and 10 mM ascorbic acid was used to prevent radiation damage. This basic buffer was supplemented with 10 mM $MgCl_2$, 150 µM $MgCl_2$, 1 M NaCl or 150 mM NaCl respectively for high or intermediate salt conditions or not supplemented with additional ions in case for the low salt measurements. The buffer exchange of the samples was performed using 10 kDa cutoff 0.5 mL Amicon filter units (14.000 × g for 25 min, repeated three times). Each sample was recorded in 10 × 3 s exposures and checked for ration damage. One full set of data for further analysis contains Holliday junction scattering profiles of a double labeled sample (AB), two single labeled samples (A label and B label), the unlabeled macromolecule (U), the gold nanocrystals (Au) and finally the buffer (Buf). The profiles are weighted, summed (AB and U) and subtracted (A and B) to calculate the gold-gold scattering interference pattern I_{Au-Au}. The resulting interference pattern were fitted using a maximum entropy algorithm to obtain the final distance distributions as described previously [69, 82, 83].

6.4.3 Preparation of Dye-labeled DNA for Single-Molecule FRET

Single stranded DNA-dye conjugates were prepared as described previously [171]. Briefly, ssDNA with 3' Amino modifier, 3' biotin modification and unmodified oligonucleotides were orderer from IDT. The sequences and modifications for all oligonucleotides are presented in Tables 6.1 and 6.2. For dye labeling the residual primary amines were removed using ethanol precipitation. After precipitation the DNA was resuspended in a 100 mM phosphate buffer, pH 8.7. As FRET pair Cy3B and Cy5 was chosen. Both NHS ester fluorophores were separately suspended in 3.5 µL DMSO and immediately afterwards 1 µL of dissolved dye was added to the corresponding aqueous oligonucleotide solution (strands H and R). The mixtures were incubated at 37 °C for 1 h and both excess dye and unreacted ssDNA were removed using polyacrylamide gel electrophoresis. Purified ssDNA-dye conjugates were extracted from the gel using squeeze and freeze and finally sample concentration was checked using a NanoDrop

device. Furthermore, 3' biotin modified oligonucleotides and unmodified ssDNA were purified using a Dionex DNAPac 200 column with a salt gradient ranging from low salt (10 mM NaCl and 20 mM sodium borate, pH 7.8) up to high salt (1.5 M NaCl and 20 mM sodium borate, pH 7.8). To assemble the full Holliday junction construct, all four complementary DNA strands (B, H, R and X) were mixed in an equimolar ratio and additionally a solution containing 10 mM Tris-HCl, 1 mM EDTA, 10 mM $MgCl_2$, pH 8.0 was added. After mixing, the solution was heated to 65 °C for 2 min followed by 39 °C for 10 min. Finally, the construct was purified using ion exchange chromatography (Dionex DNAPac 200 column) and the concentration was determined using UV-absorbance at 260 nm.

6.4.4 Single-Molecule FRET Measurements and Data Processing

Single-molecule FRET measurements were performed using a custom-built optical microscope setup and a widely used preparation protocol [172]. Flow channels for experiments were prepared as following. Firstly, channels were filled with buffer containing 10 mM or 150 µM $MgCl_2$ respectively supplemented with 30 mM Tris-HCl at pH 7.4. Secondly, 12 µL of a 1 mg/ml Biotin-bovine serum albumin (BSA) solution was added and incubated for 4 min. Thirdly, the cell was washed with a buffer containing the specified salt condition. Moreover, the coverslip was coated with streptavidin such that a final density of ~1 molecule/5 μm^2 was achieved. The streptatvidin mixture was incubated for 4 min. After incubation, the channel was washed with buffer containing the specified salt concentration again. Next, 2 pM of biotinylated Holliday junction sample was added and immobilized during the incubation for 4 min. Afterwards, unbound molecules were washed off with 20 volumes of buffer containing the specified salt condition. Finally, the channel was flushed with 2 volumes imaging buffer. Imaging buffers included the specified concentration of salt and in addition a standard oxygen scavenging system to slow down photobleaching including 2 mM protocatechuic acid (PCA), 0.001 units/µL protocatechuate-3,4-dioxygenase (PCD) and 1 mM Trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid). Data were recorded with 306 frames per second and obtained traces were analyzed using the SMART software package [173].

6.5 Supplementary Material

6.5.1 Supplementary Figures

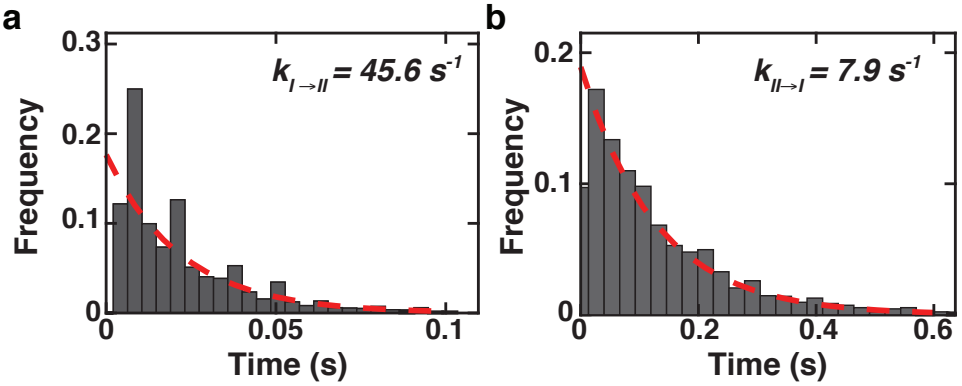

Figure 6.6: Distribution of dwell times in the low (**a**) and high (**b**) FRET states. Exponential fits are shown (red dashed lines) along with the fitted single-exponential rate constants.

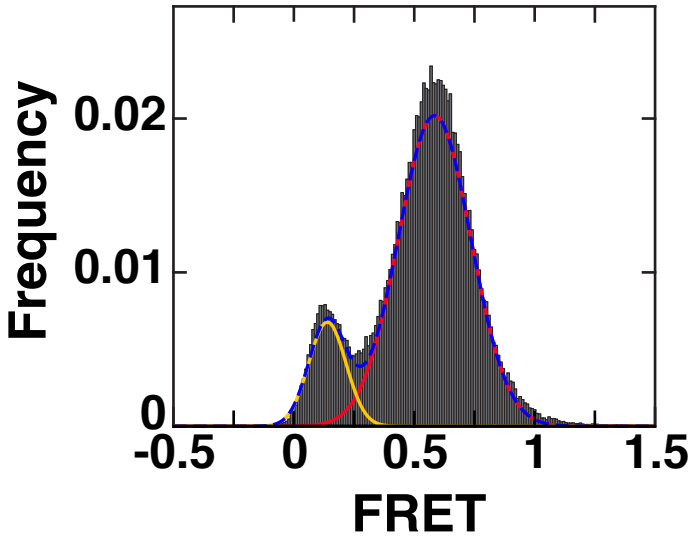

Figure 6.7: Cumulative FRET distributions. Distribution was fitted to a sum of two Gaussians (blue dashed line). The observed equilibrium constant from cumulative data was determined from the ratio of area under the high (red) and low (yellow) FRET components of the distribution.

6.5.2 Supplementary Tables

DNA Strand	Sequence
H	5’ – CCCTAGCAAGGGGCTGCTACGG
R	5’ – CCGTAGCAGCCTGAGCGGTGGG
X	5’ – CCCACCGCTCAACTCAACTGGG
B	5’ – CCCAGTTGAGTCCTTGCTAGGG

Table 6.1: Unmodified DNA sequences used to assemble the Holliday junction.

DNA Strand	Sequence
H–thiol	5’–CCCTAGCAAGGGGCTGCTACGG – thiol
R–thiol	5’–CCGTAGCAGCCTGAGCGGTGGG – thiol
X–thiol	5’–CCCACCGCTCAACTCAACTGGG – thiol
B–thiol	5’–CCCAGTTGAGTCCTTGCTAGGG – thiol
H–Cy3b	5’–CCCTAGCAAGGGGCTGCTACGG – Cy3b
R–Cy5	5’–CCGTAGCAGCCTGAGCGGTGGG – Cy5
X–biotin	5’–CCCACCGCTCAACTCAACTGGG – biotin

Table 6.2: Modified DNA sequences used to assemble the Holliday junction.

Chapter 7

Gold Nanocrystal Labels Provide a Sequence-to-3D Structure Map in SAXS Reconstructions

Summary

Small-angle X-ray scattering (SAXS) is a powerful technique to probe the structure of biological macromolecules and their complexes under virtually arbitrary solution conditions, without the need for crystallization. While it is possible to reconstruct molecular shapes from SAXS data *ab initio*, the resulting electron density maps have a resolution of ~1 nm and are often insufficient to reliably assign secondary structure elements or domains. We show that SAXS data of gold-labeled samples significantly enhance the information content of SAXS measurements, allowing the unambiguous assignment of macromolecular sequence motifs to specific locations within a SAXS structure. We first demonstrate our approach for site-specifically internally and end-labeled DNA and an RNA motif. In addition, we present a protocol for highly uniform and site-specific labeling of proteins with small (~1.4 nm diameter) gold particles and apply our method to the signaling protein calmodulin. In all cases, the position of the small gold probes can be reliably identified in low-resolution electron density maps. Enhancing low-resolution measurements by site-selective gold labeling provides an attractive approach to aid modeling of a large range of macromolecular systems.

7.1 Introduction

Small-angle X-ray scattering (SAXS) is a powerful technique to probe the structure, dynamics, and conformational transitions of biological macromolecules and their complexes in free solution [13, 43, 66]. In a typical SAXS experiment, macromolecules in solution are exposed to

This Chapter is based on a manuscript by Zettl *et al.* [3] submitted to *Science Advances* and is currently in press. I performed experiments, analyzed the data, implemented the label assignment method and contributed to writing the manuscript with help of the other authors.

an X-ray beam and the scattered photons are detected and analyzed (Fig. 7.1a). An important advantage of SAXS is that it does not require crystallization of the sample. It can probe macromolecules under a range of solution conditions, from (near) physiological to highly denaturing. This enables, for example, the detection of conformational changes as a function of pH, temperature, salt, or ligand concentration. The fact that SAXS is a solution technique has the disadvantage, however, that information is lost due to rotational averaging and that it does not provide the atomic resolution achieved by crystallography, nuclear magnetic resonance spectroscopy, and cryo-electron microscopy (cryo-EM).

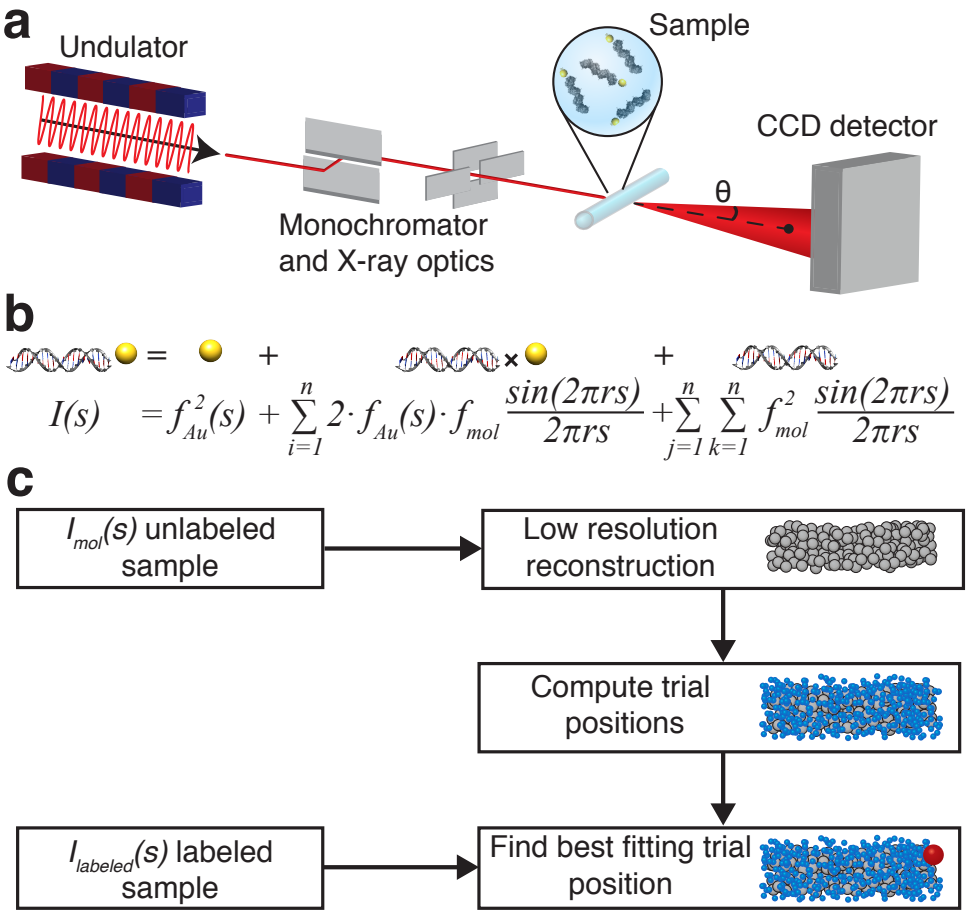

Figure 7.1: Schematic of SAXS measurements and workflow to determine gold label positions in macromolecular reconstructions. (**a**) Schematic of SAXS measurements. The incident synchrotron-generated X-ray beam (red line) is shown with X-ray optics. Single gold-labeled macromolecules are placed into the X-ray beam in a sample cell. The direct beam is blocked by a beam stop and scattered photons are detected using a CCD detector. (**b**) Scattering intensity equation for a single labeled molecule. The scattering signal can be decomposed into a sum of the individual scattering contributions: the gold label scattering, the gold-macromolecule cross-term, and the scattering from the macromolecule only. (**c**) Schematic of the workflow to determine the positions of gold-labels relative to low-resolution 3D reconstructions of unlabeled macromolecules. The SAXS profile of the unlabeled sample is used in *ab initio* reconstruction of a low-resolution model consisting of dummy residues or beads. Sterically allowed gold label trial positions are computed in the bead model. Experimental data from the gold labeled sample is compared to the computed scattering profile for each of the trial positions and the best fitting positions are identified.

Despite the comparatively limited information content, it is possible to reconstruct low-

resolution 3D electron density maps from SAXS data. Algorithms for ab initio reconstruction of low-resolution 3D structures from 1D SAXS data [14–16] have sparked an explosive growth of the use of SAXS to probe and model the structure of biological macromolecules and their complexes. The underlying principle of the reconstruction algorithms is to represent the molecule of interest as a collection of beads or “dummy residues" and to optimize the bead positions to best fit the experimental scattering profiles. The results are low (~1 nm) resolution 3D electron density maps of the samples [15, 16, 22, 174]. Unfortunately, assignment of particular domains in high-resolution protein and RNA structures to positions in a low-resolution SAXS map is often difficult, and typically requires additional information [175–181]. Fundamentally, a SAXS measurement provides information about the distribution of pairwise distances within a structure, but taken in isolation, does not reveal the specific identities of the atoms or residues that contributed to the distribution.

Recently, SAXS and anomalous SAXS (ASAXS) measurements on molecules modified by two gold nanocrystal probes have been used to measure highly accurate distance distributions between the probes [2, 41, 69, 82, 105–107, 138]. While these approaches were able to determine the spacing of the two gold-labels very precisely, a general method to relate the label position or distances to ab initio generated low-resolution maps is currently lacking. Here we demonstrate that single gold nanocrystal labels attached to specific residues of biological macromolecules can serve as fiducial markers, to provide a map from the primary sequence (and therefore domain structure) to the low-resolution 3D electron density map computed from SAXS data. We show that our approach is broadly applicable by evaluating its performance on gold-labeled DNA and RNA constructs with known structures, using a range of previously described labeling strategie [69, 105, 107]. In addition, we extend its application to gold-labeled proteins using a novel labeling approach. Our results suggest that the positions of the gold-labels can be assigned in the low-resolution map with near base-pair (bp) resolution (down to ~2 Å) and enable the assignment of protein or nucleic acid sub-domains.

7.2 Results

According to the Debye formula [66] the scattering intensity from a collection of N scatterers is given by:

$$I(s) = \sum_{i=1}^{N} \sum_{j=1}^{N} f_i(s) \cdot f_j(s) \frac{\sin(2 \cdot \pi \cdot s \cdot r_{ij})}{2 \cdot \pi \cdot s \cdot r_{ij}} \tag{7.1}$$

were f_i and f_j denote the individual scattering factors of scatterers i and j, s is the magnitude of the momentum transfer (with $s = 2\sin(\theta)/\lambda$, where $2 \cdot \theta$ is the total scattering angle and λ the X-ray wavelength) and r_{ij} the distance between the i^{th} and j^{th} scatterer. For a gold-labeled macromolecule, the scattering profile $I(s)$ contains contributions from the gold nanocrystal label, from pairs of scatterers in the macromolecule, and from pairs consisting of one macromolecular scatterer and the gold label (Fig. 7.1b). Therefore, the overall scattering profile depends on the position of the gold probes relative to the macromolecule (Fig. 7.1b).

An overview of our approach is as follows: We directly measure the scattering profiles of the bare gold probes $I_{Au}(s)$, the unlabeled macromolecule $I_{mol}(s)$, and the macromolecule with a single attached gold probe $I_L(s)$. We compute the profile for cross-scattering between the

gold probe and the macromolecule, $I_{Au-mol}(s)$, by subtracting the gold probe profile and the unlabeled macromolecule profile from the profile of the singly labeled macromolecule, after normalization. The scattering profile of the unlabeled macromolecule was used to generate an *ab initio* low-resolution 3D electron density map. The low-resolution bead model was then used to generate a set of sterically allowed trial gold positions. The experimentally measured cross-scattering term, $I_{Au-mol}(s)$, was then compared to the predicted cross-scattering terms for each of the trial gold label positions. The trial position that gives the best fit was identified as the most likely gold-label placement. A schematic of our workflow is presented in Fig. 7.1c.

We employed 7 Å radius thioglucose-passivated gold nanocrystals synthesized by the Brust method (see Methods 7.4 as fiducial markers attached to a range of different biological macromolecules. As test samples, we used *i*) double-stranded DNA (dsDNA) molecules ranging from 10 to 35 bp length labeled at their ends; *ii*) dsDNA molecules of a fixed length of 26 bp internally labeled at various positions; *iii*) kinked-turn RNA constructs, labeled at two respective ends of the RNA motif; and *iv*) the signaling protein calmodulin labeled at two different positions, corresponding to the two lobes of its known structure (see Materials and Methods for details of the sample preparation and labeling procedures 7.4). SAXS data were record at beamline 12-ID of the Advanced Photon Source (APS) and at beamline 4-2 of the Stanford Synchrotron Radiation Lightsource (SSRL) (Fig. 7.1a and Methods 7.4). Since matched sample concentrations are needed in the analysis but challenging to achieve experimentally [69], intensity profiles were corrected computationally (see Methods 7.4). In a first step, $I_L(s)$ was normalized by matching its tail to the tail of $I_{Au}(s)$ at $s \geq 0.04$ Å^{-1} (see Methods for details 7.4), where the scattering is dominated by the gold probe (compare Fig. 7.2a and b, yellow lines and red lines). In a second step, Guinier analysis was applied to estimate the forward scattering intensity for all scattering profiles (Fig. 7.2a and b) and used to normalize the molecule only scattering profile $I_{mol}(s)$.

Next, we analyzed the gold-only scattering I_{Au} by decomposing the experimental scattering signal into a linear combination of basis profiles for spherical scatterers [69] (Fig. 7.7). Non-negative least square fits showed excellent agreement with the experimental data (Fig. 7.8) and suggested a narrow size distribution of the employed gold nanocrystals. The fitted representation of the gold scattering was used subsequently as the form factor $f_{Au}(s)$ for the gold particles in evaluation of the Debye formula Eq. 7.1.

We built *ab initio* low-resolution models of the macromolecules under study using the scattering data from the unlabeled macromolecules using established procedures (Methods 7.4). The low-resolution reconstructed electron density maps show good agreement with the high-resolution structures for all measured DNA constructs assuming standard B-form DNA helix geometries [58] (Figs. 7.2c, 7.4b), the RNA kink-turn motif [107, 168, 182] (Fig. 7.5b), and the calcium-bound holo-conformation of calmodulin [73, 183] (Fig. 7.2d).

To determine the gold label position, we created gold marker trial positions by randomly generating ~100 positions on an 11 Å radius sphere (corresponding to the extension of the 7 Å radius gold particle and the ~4 Å linker) around each bead in the low-resolution reconstruction (see Methods 6.4). Subsequently, positions that result in steric clashes were eliminated. The sterically allowed trial positions (typically ~1000-25,000 for the structures investigated here) enveloped the entire reconstructed low-resolution maps (Fig. 7.2e and f). We then computed the total scattering intensity for every gold-label trial position by adding the gold label and unlabeled macromolecule scattering terms to the calculated scattering cross-term between the

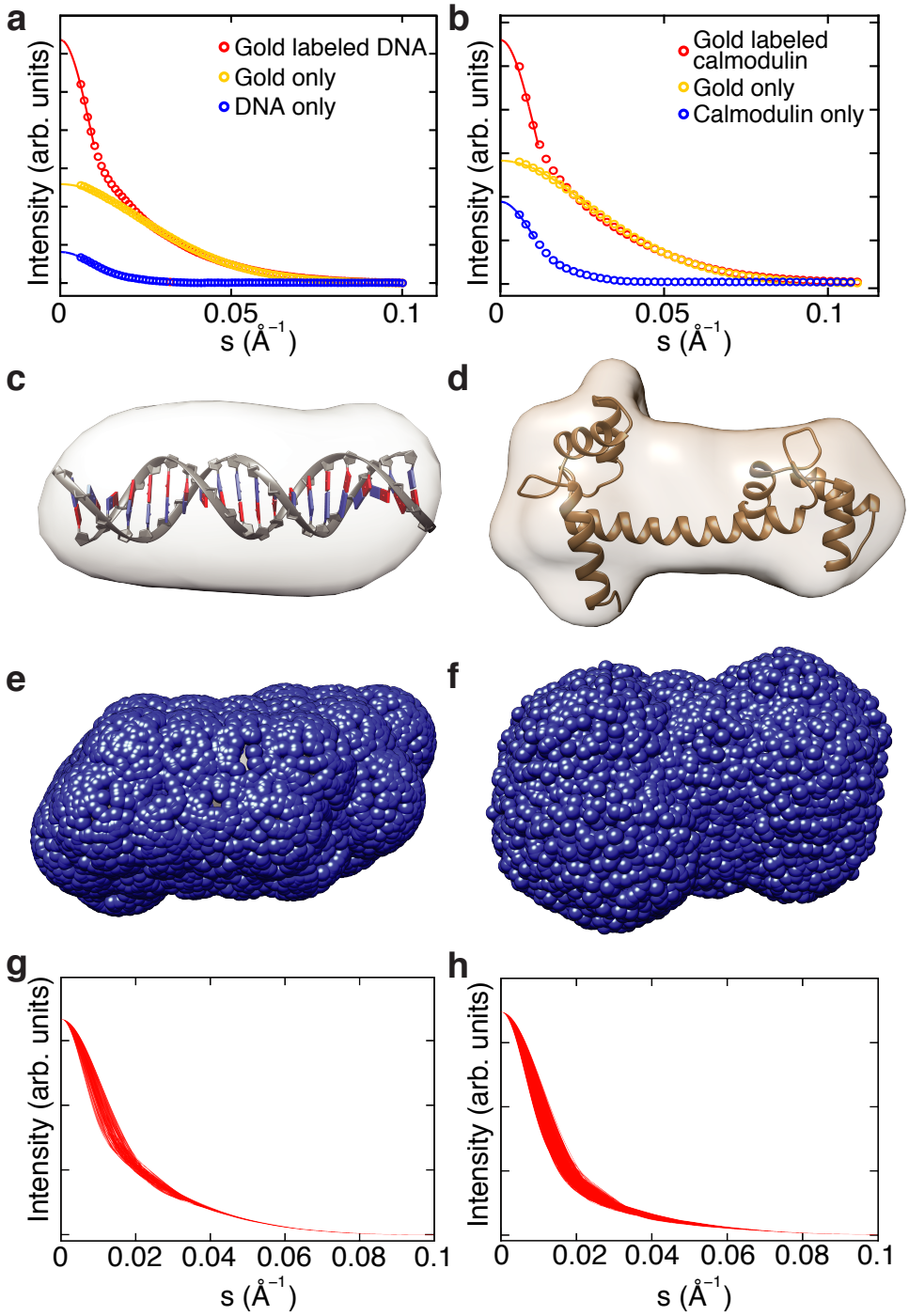

Figure 7.2: SAXS shape reconstruction, trial gold label positions, and computed scattering intensity profiles. (**a**), (**b**) SAXS profiles and Guinier fits for representative samples (20 bp end-labeled DNA in panel (a) and calmodulin labeled at position N111C in panel (b). Experimental SAXS profiles are shown as circles. Guinier fits to determine $I(0)$ are shown as solid lines. (**c**), (**d**) 3-D reconstructions of 20 bp DNA in (c) and calmodulin in (d). The known high-resolution structures are superimposed [73]. (**e**), (**f**) Trial gold label positions (blue spheres) calculated around the envelope of the 3-D reconstruction for the 20 bp DNA (e) and calmodulin (f). (**g**), (**h**) Calculated total intensity profiles for a reduced set of trial gold label positions for 20 bp DNA (g) and calmodulin (h). All trial profiles were ordered from the best to the worst fitting and every 20$^{\text{th}}$ curve is shown for clarity, resulting in 91 traces for 20 bp DNA (g) and 1200 traces for calmodulin (h).

gold marker and macromolecule. We found that different trial positions resulted in significantly different calculated scattering intensities $I_{total}(s)$, with the resulting profiles depending on the relative label-molecule placement (Fig. 7.2g and h), suggesting that the scattering profiles of the gold-labeled macromolecules contain the desired information about the relative gold probe-macromolecule arrangement. To determine the best trial position, we evaluated the normalized square difference between the computed profiles and the experimental scattering profile of the labeled macromolecules (Eq. 7.13 in Methods 7.4). The trial positions were ordered from the best to worst fitting. We associated the best fitting trial position with the most likely gold marker placement.

7.2.1 End-labeled dsDNA constructs

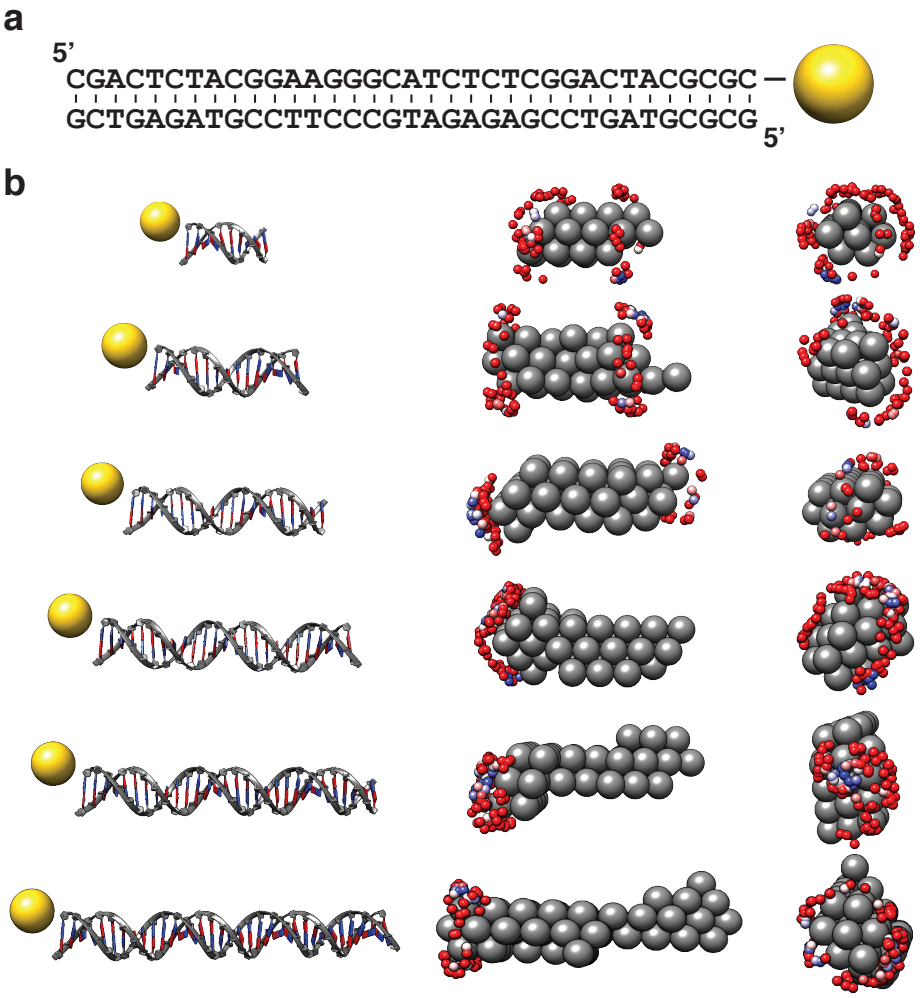

Figure 7.3: Determination of gold-label positions for end-labeled DNA. (**a**) Secondary structure of the 35 bp DNA duplex and the gold label position indicated on the 3' end of the upper strand. (**b**) Mapping gold label positions to 3D reconstructions using SAXS data for end-labeled DNA. Left column: Schematic representations of the 10, 15, 20, 25, 30, 35 bp DNA constructs end labeled with gold nanocrystals used in the study. Middle and right columns: Side (middle) and front (right) views of the 3D reconstructions of the unlabeled samples (shown as grey spheres) and mapped positions of the gold labels (blue to red spheres; blue spheres correspond to the best fit). The placement of the gold labels at the end and off center is well reproduced.

We first applied our procedure to determine the gold-label positions to end-labeled dsDNA molecules ranging from 10 bp up to 35 bp (Table 7.1). The low-resolution reconstructed structures exhibited cylindrical shapes with a width and length as expected for our different DNA constructs (Fig. 7.3b). The calculated profiles for the gold form factor model, the 3D bead reconstruction, and the best fitting trial position showed excellent agreement with the experimental data for the gold nanocrystals, the unlabeled DNA molecules, and the labeled molecules, respectively (Fig. 7.8).

All trial gold label positions were ordered from best to worst fitting for each DNA length. For visualization, the best 100 positions (out of 1055 to 2521 total) are shown, with the top 10 positions colored from blue to white and the next 90 positions colored from white to red. The best fitting positions were located at the ends of the reconstructed cylindrical density maps for all investigated DNA lengths and off the cylindrical axis (Fig. 7.3b), as expected for the end-labeled constructs with a linker that positions the gold nanocrystal label off-axis [2, 82]. The DNA constructs are almost symmetric under both rotation along the helix axis and exchange of the ends. These symmetries are visible in the reconstructed label positions: at each end, the best fitting positions resulted in a doughnut-like shape (Fig. 7.3b, right panel). In addition, we observed well-fitting gold label positions at both ends of the reconstructed cylindrical shapes.

Since the scattering contribution (in particular the forward scattering) increases approximately quadratically with molecular mass [43, 66], the contribution of the DNA relative to the (fixed-size) gold label increases with increasing DNA length for the gold-labeled DNA constructs (Fig. 7.8; compare the square and circle data points). The larger relative signal from the DNA appears to enhance the positioning accuracy in our gold-label assignment procedure, as evidenced by a reduced scatter and stronger clustering of the best fitting trial positions for the longer DNA constructs (Fig. 7.3b). In summary, the data for end-labeled DNA construct demonstrate that it is possible to reliably assign the label positions to the end of the helix and off-axis for all investigated DNA lengths.

7.2.2 Internally labeled dsDNA

To test our method for labels attached in the interior of a nucleic acid sequence, we used dsDNA-gold constructs with a fixed length of 26 bp and varying internal label positions (Fig. 7.4a, Table 7.2). Again, the different calculated profiles gave excellent fits to the experimental scattering profiles of the single-labeled DNA, the unlabeled DNA, and the gold nanocrystals (Fig. 7.9). The data clearly indicate that distinct label positions yield different cross-scattering terms (Fig. 7.9, red lines) and thus can be distinguished. All trial label positions were again ordered from best to worst fitting and the best fitting positions for each of the internal label placements are shown in Fig. 7.4b. The nomenclature for the internal label positions is based on the distance of the labeled base position (in bases) and the DNA strand (strand A or B, Table 7.2), counting from the 5' ends. For example, A4 (Fig. 7.4b, purple sphere) carries a label on the 4^{th} base from the 5'-end of DNA strand A and B11 (Fig. 7.4b, blue sphere) on the 11^{th} base of DNA strand B. For example, as expected the A4 (Fig. 7.4b, purple sphere) label position is located near the end of the reconstructed DNA molecule (Fig. 7.4b, grey envelope), whereas the B11 (Fig. 7.4b, blue sphere) position was determined to be almost at the midpoint of the duplex.

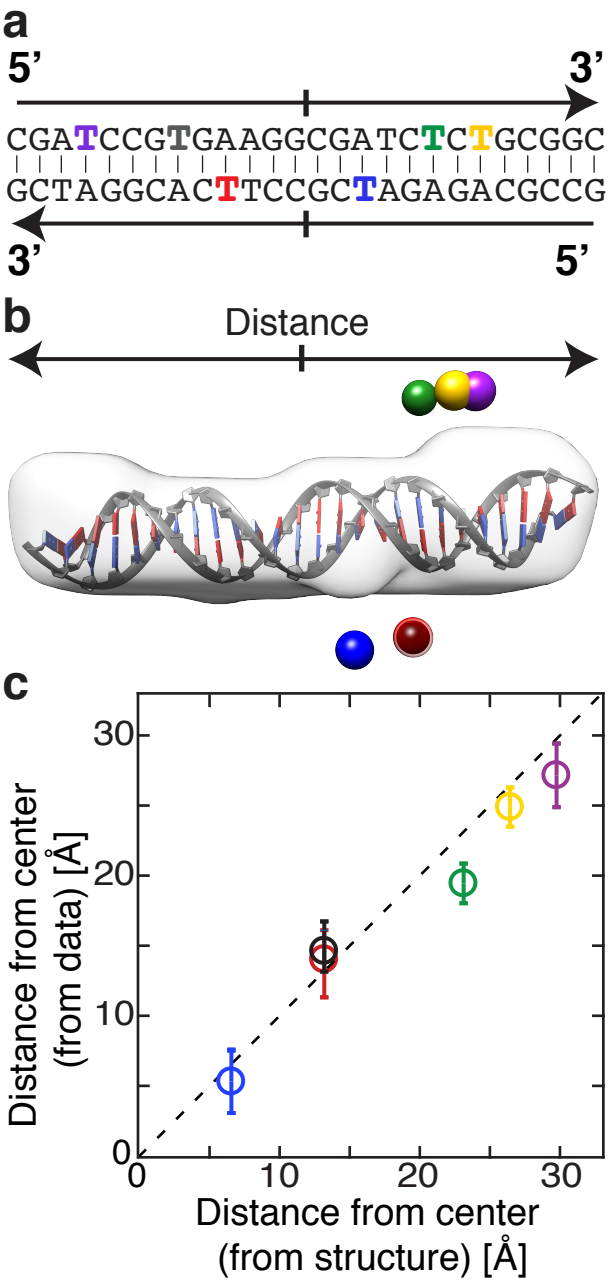

Figure 7.4: Determination of gold-label positions for internally labeled DNA. (**a**) Secondary structure of the 26 bp DNA duplex. Colored letters indicate bases used for internal labeling. The same color code is used in (b) and (c). (**b**) Low-resolution shape reconstructions of the unlabeled DNA (grey) superimposed on a high-resolution structure. The best fitting label positions for the six different internal gold-label positions are shown as colored spheres. (**c**) Distance to the center of the DNA computed as the mean and standard deviation of the best 50 reconstructed and filtered label positions for each attachment point as a function of the distance of the labeled base from the center of the helix (converted using a helical rise per base [69, 82] of 3.3 Å). The dashed line indicates 45°. The RMSD between computed and expected positions is 2.1 Å.

We computed the mean and standard deviation of the distance from the center of the reconstructed shape along the helical axis (indicated by the arrows in Fig. 7.4b) for the 50 best fitting positions for all six internally labeled constructs. We note, that similar to the end-labeled DNA constructs discussed in the previous section, the very high degree of symmetry of the bare DNA duplex does not allow us to uniquely assign the ends. We can, however, determine the distance from the center of the duplex with high accuracy for each of the label positions: the reconstructed label positions are compared to the distance from the center that is expected from the DNA structure taking into account that the label position is shifted by slightly less than one base in the 3'-direction due to the orientation of the attachment to the DNA [105] (Fig. 7.4b). Overall, we find excellent agreement between the computed and expected label positions (the root-mean-square deviation is RMSD = 2.1 Å and the reduced χ^2 = 1.7; Fig. 7.4c). Label positions A8 and B17 are expected to have almost identical distances to the center, which is well reproduced by the best fitting trial positions (Fig. 7.4). We can reliably distinguish the position of labels separated by only 1-2 base pairs (Fig. 7.4c), corresponding to a distance of ~3-5 Å. The average deviation between measured and expected positions is 0.6 bp or 1.9 Å. Taken together, our results suggest that our procedure to assign the position of gold-labels can achieve a few Ångström resolution, close to the resolution achieved in X-ray scattering interferometry measurements with two gold-labels [2, 69, 105–108] and better than the ~10 Å resolution for standard SAXS measurements.

7.2.3 RNA kink-turn motif

Having demonstrated applicability of our method to DNA, we next tested it on gold-labeled RNA constructs. We used a RNA kink-turn motif (Table 7.3) that was modified with gold nanoparticles at the 3' end positions [83, 107] (Fig. 7.5a). RNA kink-turn motifs are commonly found in functional RNA's, including the ribosome. They consist of a three-nucleotide bulge flanked by a GA/AG tandem base pair, which stabilizes a kink of more than 90 degrees in the flanking helices [107, 168, 182]. Again, *ab initio* shape reconstruction was used first to compute a low-resolution electron density map, which showed good agreement with a high-resolution model [107] (Fig. 7.5b). The experimentally recorded and reconstructed profiles were again in excellent agreement for the best fitting trial positions at both probe locations (Fig. 7.10). The trial positions were filtered (see Methods 7.4) and the final calculated probe positions for the two gold labels were located on opposite sides of the RNA kink-turn motif (Fig. 7.5b, c). The computed gold label positions (Fig. 7.5b, blue and red spheres) show excellent agreement with the positions of the probe-modified bases in the high-resolution model. For further analysis and quantification, the low-resolution density map and the calculated trial points were aligned (Fig. 7.5b). The positions of the trial points along the x-axis was defined as in Fig. 7.5c, and their distances from the geometrical center of the reconstructed shape were calculated (Fig. 7.5b and c, grey shape). We note that the RNA kink-turn motif in our study is almost symmetric, having 12 nucleotide pairs in the "arms" on either side of the 3-nt (nucleotide) central kink-turn bulge. The only difference between the two arms is that one side is fully base-paired whereas the other side has a three-base mismatch next to the central bulge (Fig. 7.5d, Table 7.3). The clear separation of the fitted label position for the labels at the two distinct termini demonstrates that specific nucleotide sequences can be unambiguously located in low-resolution reconstructed density maps, even for relatively minor differences in the structures.

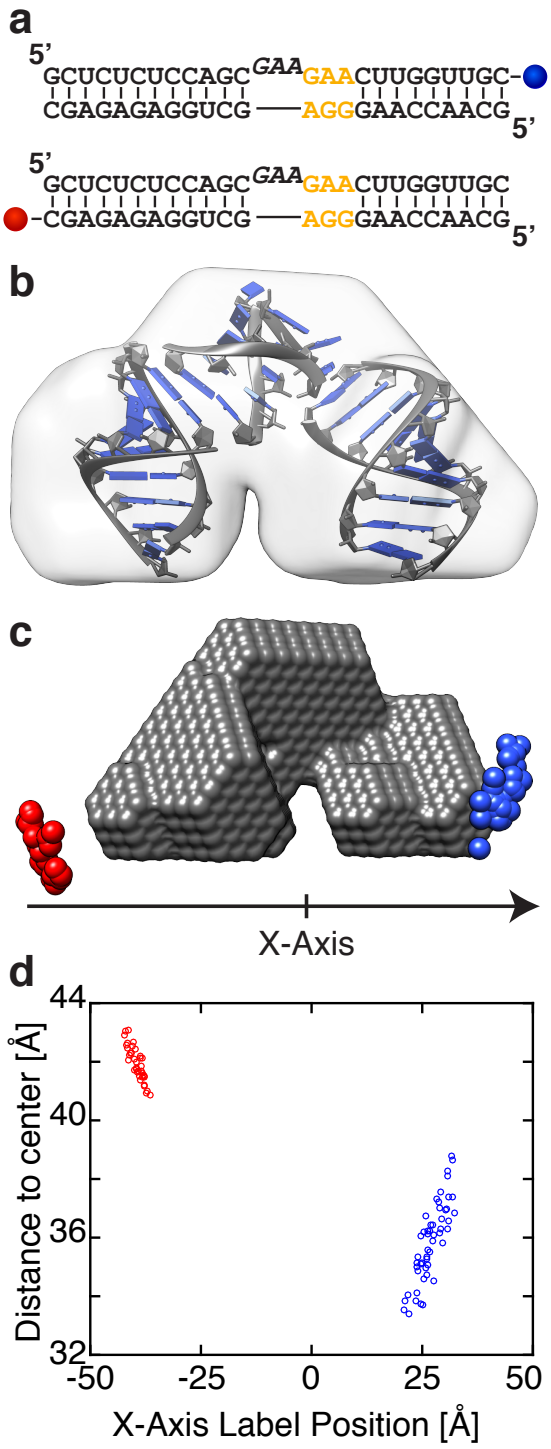

Figure 7.5: Gold-label positions for an RNA kink-turn motif. (**a**) Secondary structure of the RNA kink-turn and the two labeling position at the 3' ends for the individual constructs. The 3-nt kink-turn bulge is displayed with a small vertical offset and the 3-base mismatch region is shown in orange. (**b**) Shape reconstruction of the unlabeled RNA kink-turn (light grey shape) superimposed with the high-resolution structure [107]. (**c**) Low-resolution model (grey spheres) and best fitting gold nanocrystal positions (blue and red spheres) after filtering. (**d**) Scatter plot of filtered gold label positions. The X-axis value of the fitted label positions is plotted vs. the overall distance to the geometrical center of the low-resolution shape reconstruction.

7.2.4 Gold-labeled calmodulin constructs

Having demonstrated our method for a range of DNA and RNA constructs that relied on proven labeling chemistries, we next developed a protocol to attach the same gold nanocrystal labels to proteins. The ability to label proteins extends our approach to the third major class of biological macromolecules. We employed two different mutants of the calcium-binding messenger protein calmodulin that carry single cysteine mutations at selected positions (R37C and N111C, see Methods 7.4) as a model system. We took advantage of the thiol side chain of cysteine to couple gold nanocrystals directly to the selected positions of the protein [111–113]. We focused on the calcium bound state, i.e. the holo conformation of calmodulin, which features two roughly globular and distinct lobes [73].

While calmodulin is well known to undergo large conformational changes upon Ca^{2+} binding or interactions with other binding partners, under the conditions of our experiment ($\geq$5 mM Ca^{2+}; measurement with 5 and 10 mM Ca^{2+} give identical results, within error; Fig. 7.11a) calmodulin has two rigid globular Ca^{2+}-binding domains connected by an α-helical linker region that provides interdomain flexibility [184–190]. The assumption of an overall folded protein with two lobes that are connected by a linker region with some flexibility is supported by several lines of evidence: *i*) the Kratky representation [$s^2 \cdot I(s)$ vs. s] of the scattering data is indicative of a folded protein (Fig. 7.11b); *ii*) the histogram of pairwise distance $P(r)$ is consistent with the typical bend dumbbell shape of calmodulin in the presence of Ca^{2+} (Fig. 7.12f) [190–192]; *iii*) our experimental SAXS profiles agree reasonably well with the predicted profile [193] from an available crystal structure of Ca^{2+}-bound calmodulin (Fig. 7.11c), in line with previous SAXS work [192], but the shoulder at $s \approx 0.025$ Å^{-1} is less pronounced in the experimental data, indicating a smaller separation between the two domains than in the crystal structure; and *iv*) multiple *ab initio* reconstruction runs converged on onto a single structure as indicated by pair-wise normalized spatial discrepancy [72] values of $\approx$0.5, in good agreement with previous SAXS work that has modeled calmodulin with a single conformation under similar experimental conditions [192].

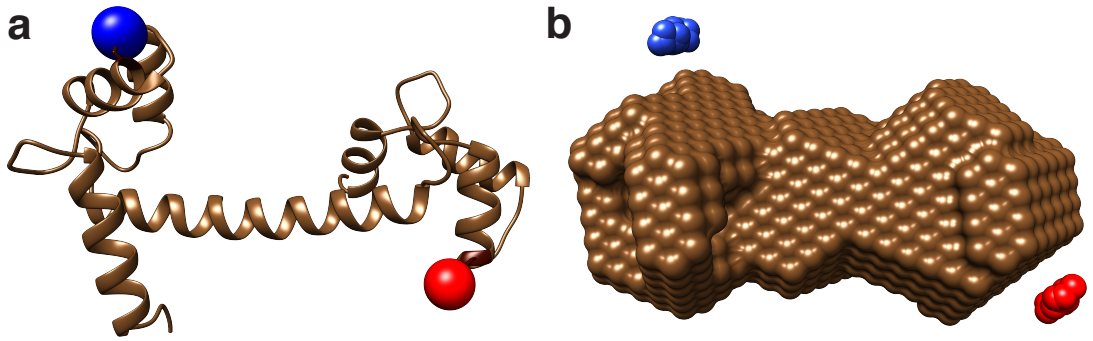

Figure 7.6: Gold-label position reconstructions for labeled calmodulin constructs. (**a**) Crystal structure of Ca^{2+}-bound calmodulin [73] with mutated amino acid positions indicated by colored spheres (R37 blue and N111 red). (**b**) View of the 3D reconstructions of the unlabeled protein (shown as brown spheres) and mapped positions of the gold labels at R37C (blue) and N111C (red) as colored spheres for the filtered final label positions. The placement of the gold labels at the designated positions is well reproduced.

The two mutants were chosen such that the cysteines are positioned at the surface of the two different lobes and placed to avoid steric clashes of the gold labels with the protein.

Labeling of similar calmodulin constructs with cyanine dyes [194] or large gold nanoparticles (~40-60 nm diameter via a biotin-streptavidin linkage) [195] left them fully functional. In addition, attachment of ~1.5 nm gold nanoparticles coated with uncharged ligands, similar to the particles used in this work, directly to cysteine residues of cytochrome *c* did not perturb the structure of the folded protein [111].

For data analysis and gold marker position reconstruction of the labeled calmodulin data, we used the same procedure as for the labeled nucleic acid constructs. For each construct, the best fitting trial positions were calculated and sorted according to the quality of fit. The best trial positions were filtered as described in Methods (7.4). The reconstructed intensity profiles show very good agreement with the experimentally measured profiles (Fig. 7.13). The best fitting trial positions after filtering for the R37C (Fig. 7.6a, green spheres) and N111C constructs (Fig. 7.6a, red spheres) match the positions of these amino acids in the high-resolution structures (Fig. 7.6a, b). The Ca^{2+}-bound calmodulin has two approximately symmetric helix-loop-helix domains separated by linker [73]); the clear separation of the reconstructed gold-label positions suggests that our method can assign the accurate positions of both labels despite the flexibility of the central linker and, therefore, distinguish domains with even relatively small differences in structure.

7.3 Discussion

We have used SAXS measurements to determine gold nanocrystals label positions relative to low-resolution electron density maps computed from SAXS data for a range of biological macromolecules, including 10-35 bp long end-labeled dsDNA, internally labeled dsDNA, an RNA kink-turn motif, and calmodulin as a protein model system. Combined, our results show that our method is broadly applicable to all major classes of biological macromolecules and that by using small (~7 Å radius, ~80 gold atoms) gold nanocrystals as fiducial markers, the position of gold probes and the corresponding labeled residues can be accurately located in low-resolution reconstructions. Determining the 3D position of specific labeled residues allows the sequence of the macromolecule to be placed within the macromolecular shape. The gold labeling approach clearly distinguishes different domains and secondary structure elements.

Our work opens up several exciting possibilities and future directions. The demonstrated ability to site-specifically label proteins with small gold nanocrystals creates the possibility to extend X-ray scattering interferometry gold-gold distance measurements by SAXS [69, 82, 83] or ASAXS [2] to proteins and analysis of the structural arrangement of protein-nucleic acid complexes. Application of our reconstruction method to doubly gold-labeled macromolecules could provide additional information, in particular to resolve ambiguities in the relative placements of residues, for example in reconstructions that have internal symmetry. Our current protocol assumes that the macromolecule of interest adopts a well-defined structure or at least one dominant conformation in solution. It might be possible to relax this assumption in the future, since SAXS is capable, in principle, to probe conformational ensembles [196–198]. While it has been shown that labeling with small gold nanoparticles with uncharged ligands tends to maintain the structure of protein and nucleic acids, for macromolecules or complexes of unknown structure, in general, testing the functionality and structural integrity of the labeled constructs will be important.

While this proof-of-concept study uses gold nanoparticles as labels, we expect our method

to be equally applicable to other labels, such as silver and platinum nanoparticles, which could provide equally attractive and orthogonal labeling options and might permit simultaneous measurements of several marker positions using ASAXS. Finally, we envision that our general approach will not be limited to SAXS. Zhang and coworkers recently demonstrated that gold labels can be used to obtain structural and dynamical information for labeled DNA duplexes in cryo-EM measurements [199]. In conclusion, nanometer sized metal probes provide a powerful new approach to determining precise residue positions in biological macromolecule. We anticipate that this approach will provide new and quantitative insights into macromolecular structure, dynamics, and interactions.

7.4 Materials and Methods

7.4.1 Gold nanoparticle preparation

Water-soluble gold nanocrystals were synthesized and purified as described previously [69, 82, 83]. The resulting gold nanocrystals are passivated with a 1-thio-β-D-glucose ligand shell to stabilize the gold nanoparticles and to achieve a highly monodisperse size distribution [200] (Fig. 7.7).

7.4.2 Preparation of gold-labeled DNA and RNA samples

All DNA/RNA sequences used in this study are given in Tables 1-3 of the Supporting Material (Tables 7.1, 7.2, and 7.3). The DNA and RNA oligonucleotides for gold-labeling at the ends were prepared on an automated ABI 394 DNA synthesizer (Applied Biosystems) and thiols were incorporated using a C3-thiol-modifier (Glen Research, part # 20-2933-41) on the 3'-end of the single-stranded DNA or RNA. Oligonucleotides were purified using ion-exchange high-pressure liquid chromatography (HPLC) followed by reversed-phase HPLC. The coupling of the gold nanoparticles to the and 3' ends of DNA was as described by Mathew-Fenn et al. [69, 82] and the protocol for gold-labeling RNA constructs was carried out as described by Shi et al. [108]. Briefly, a fivefold excess of gold nanocrystals was mixed with the thiol-modified single-stranded DNA or RNA molecules in a buffer solution at a pH of 9.0. After two hours, uncoupled gold nanoparticles or gold nanocrystals conjugated with multiple ssDNA strands were removed by ion-exchange HPLC and the product was stored at −20°C. A NanoDrop ND-1000 (NanoDrop Technologies) was used to determine final sample concentrations by quantifying the absorbance at 260 nm and 360 nm. The single stranded DNA oligonucleotides were annealed with their complementary strands at room temperature for 30 min whereas RNA oligonucleotides were annealed at 40°C for 30 min.

The procedure for preparing the internally labeled DNA constructs was as described [105]. The oligonucleotides (Table 7.2) with internal amino-modified thymine (Amino-Modifier, C2 dT) were prepared on an ABI 394 DNA synthesizer (Glen Research, part # 20-2933-41) and reacted with a solution containing 1 mg SPDP [succinimidyl 3-(2-pyridyldithio)propionate] per 10 µL DMSO and the disulfide bridge was cleaved using a 200 mM DTT, 50 mM Tris-HCl, pH 9.0 solution afterwards. Finally, complementary DNA strands were hybridized and purified by anion exchange HPLC to obtain pure single labeled samples.

7.4.3 Preparation of gold-labeled protein samples

We used calmodulin as a protein model system. Calmodulin is a well-characterized and stable protein with a known crystal structure; in addition, it has the advantage that the wild type is naturally cysteine-free. Calmodulin was expressed and labeled with gold nanocrystals as follows: A single cysteine was introduced in sea urchin calmodulin by means of the mutations R37C or N111C. This calmodulin was expressed in *Escherichia coli* and purified as described by Gopalakrishna and Anderson [201]. The calmodulin [800 µM in 50 mM Tris-HCl, pH 7.4, 1 mM DTT] was incubated in 150 µL of 10 mM DTT, 50 mM Tris-HCl, pH 9.0. Excess DTT was removed by fast-flow G25-sepharose spin filtration (GE Healthcare) and immediately after purification a 5-to-1 ratio of purified and desalted Au nanocrystals to protein and 20 µL of 1 M Tris-HCL, pH 9.0 was added to the solution and incubated for 2 h at room temperature. The reaction was stopped by adding 15 µL 2 M ammonium acetate solution, pH 5.6. Finally, unreacted Au nanocrystals and unlabeled proteins were removed using the Superdex 75 gel filtration column.

7.4.4 SAXS measurements

Small-angle X-ray scattering experiments were performed at the BESSRC-CAT beamline [202] 12-ID of the Advanced Photon Source (APS; for end-labeled DNA and calmodulin) and at beamline 4-2 of the Stanford Synchrotron Radiation Lightsource (SSRL; for end-labeled and internally-labeled DNA and the RNA kink-turn). At beamline 12-ID, we employed an X-ray energy of 12 keV, a sample detector distance of 1 m, and a custom-made sample cell [19]. For calibration, a silver behenate standard was used to locate the beam center and calibrate the scattering angle. An exposure time of 10×1 s and a CCD detector was used for data collection. Data were reduced using the Goldcontrol software package. At SSRL an X-ray energy of 9 keV and a sample-detector distance of 1.5 m or 11 keV and 1.1 m sample-detector distance was chosen. Data was collected using a linear position sensitive proportional counter (LPSPC) in ten one-minute exposures and a Mar CCD165 (MAR) in twenty fifteen-second exposures. Data collected using the LPSPC detector was reduced with the OTOKO and SAPOKO software packages and data collected using the MAR detector was reduced with the Blue Ice software package and additional scripts provided at the beamline.

All SAXS measurements for DNA samples were performed in 70 mM Tris-HCl buffer, pH 8.0, with 100 mM NaCl and 10 mM sodium ascorbate added. SAXS measurements for RNA kink-turn samples were obtain in 70 mM Tris-HCl buffer, pH 7.4, with 60 mM NaCl, 10 mM $MgCl_2$, and 10 mM sodium ascorbate added and measurement for calmodulin samples used 60 mM Tris-HCl buffer, pH 8.0, with 100 mM NaCl, 10 mM sodium ascorbate, and 5 mM $CaCl_2$ unless otherwise indicated. Samples were measured at room temperature in sample cells with a 2 mm path length and 25 µm mica windows [19]. Buffer profiles were measured before and after each sample using identical procedures and subtracted for background correction. Scattering curves containing dsDNA were recorded at 50 µM and dsRNA at 30 µM sample concentration whereas calmodulin data was obtained at 250 µM.

7.4.5 SAXS data analysis

We employed SAXS3D [16] and DAMMIN [15, 21] to reconstruct the low-resolution bead models from the scattering profiles of the unlabeled samples. In all cases, repeated DAMMIN runs converged and yielded structures with pair-wise normalized spatial discrepancy [72] values <1.0. Custom written MATLAB scripts were used to fit the intensity profile of the unlabeled sample using the reconstructed bead model, the gold only scattering profile, and to reconstruct the macromolecule-gold label interference term. These intensity patterns were used to reconstruct position dependent scattering profiles of gold label-molecule constructs as described below. In addition, radii of gyration and $P(r)$ functions were computed for all samples using PRIMUS [60] and are presented in Table 7.4 and Fig. 7.12.

7.4.6 Shape reconstruction of the unlabeled sample

Ab initio shape reconstructions are based on representing the structure in solution by a collection of beads or dummy residues using the Debye formula [16] or approximations to it [15, 66]. The shape reconstruction algorithms use either simulated annealing (in the case of DAMMIN) or a “give’n’take" protocol (SAXS3D) to obtain the best fit to experimental data. In the reconstructions, the beads are treated as point scatterers, such that all f_i are identical and independent of s. Therefore, the scattering profile from a particular bead model is given by:

$$I_{mol}^{model}(s) = c_1 \cdot \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\sin(2 \cdot \pi \cdot s \cdot r_{ij})}{2 \cdot \pi \cdot s \cdot r_{ij}} + c_2 \tag{7.2}$$

The scattering pattern computed from the Debye formula is adjusted using a multiplicative constant c_1 and an additive constant c_2 to achieve the best fit to the experimental data $I_{mol}(s)$. Shape reconstructions for the 10-35 bp DNA constructs were performed using SAXS3D; control calculations with DAMMIN gave similar results. The experimental scattering profiles of the unlabeled samples were given as input. Reconstructions calculated on a 10 Å lattice achieved the most robust results for all six DNA lengths compared to smaller lattice spacing without losing to much resolution and were finally exported as PDB-coordinate files. For the 26 bp DNA, the RNA kink-turn, and calmodulin DAMMIN was used for reconstructions, with typical dummy atom radii between 1.5 to 2 Å.

7.4.7 Representation of the gold nanocrystal scattering

The gold nanocrystal labels employed in this study exhibit small polydispersity and their scattering contribution has been previously shown to be well described as superposition of spheres with different radii [69, 82]. The gold nanocrystal profile I_{Au} was fitted using a set of basis functions and coefficients determined by a non-negative least squares fit using custom-routines in MATLAB. The basis profiles $I(s, R)$ were generated assuming spherical scatterers with radii R ranging from 2 to 100 Ångström:

$$I(s, R) = 3 \cdot \frac{\sin(2 \cdot \pi \cdot s \cdot R) - 2 \cdot \pi \cdot s \cdot R \cdot \cos(2 \cdot \pi \cdot s \cdot R)}{(2 \cdot \pi \cdot s \cdot R)^3} \tag{7.3}$$

A typical radius distribution is shown in Fig. 7.7. We note that the atomic sub-structure of the particles only affects scattering at much higher s-values than used in this study [203].

7.4.8 Normalization of scattering profiles

Our analysis of the labeled samples requires appropriate normalization of the scattering intensities for the labeled, unlabeled, and gold only samples. In principle, such normalization is not necessary if all three samples are measured at identical (molar) concentrations. However, in practice it is difficult to prepare different samples at precisely matched concentrations and therefore the following normalization procedure was implemented. First, the intensity profile of the gold nanocrystal labeled sample $I_L(s)$ was matched to the intensity profile of the gold nanoclusters $I_{Au}(s)$ in such a way that the best overlap of both scattering profiles was achieved for intensities at $s > 0.04$ Å^{-1} where the scattering of the labeled sample is dominated by the gold nanocrystal (Fig. 7.2a and b) [69, 82]. Second, Guinier analysis was carried out for all three profiles to determine the forward scattering intensities $I_{Au}(0)$, $I_{L,N}(0)$ and $I_{mol}(0)$ using Primus [60], where $I_{L,N}(s)$ denotes the gold labeled profile after the initial normalization to the gold only profile. For the forward scattering intensity, i.e. for the limit $s \to 0$, the Debye formula simplifies to

$$I(0) = N^2 f_i(0) \cdot f_j(0) = N^2 \cdot f_i^2 = A^2(0) \tag{7.4}$$

where we have assumed identical f_i and defined the scattering amplitude A in the last step.

The scattering intensity of the labeled sample is the sum of the normalized scattering intensity of the gold crystal, normalized unlabeled molecule, and the scattering cross term of gold crystal and unlabeled molecule.

$$\begin{aligned} I_{L,N}(0) &= I_{Au}(0) + I_{mol}(0) + I_{Au-mol}(0) \\ &= A^2_{Au}(0) + A^2_{mol,1}(0) + A^2_{Au-mol}(0) \\ &= A^2_{Au}(0) + A^2_{mol,1}(0) + 2 \cdot A_{Au}(0) \cdot A_{mol,1}(0) \end{aligned} \tag{7.5}$$

$$A^2_{mol,1}(0) + 2 \cdot A_{Au}(0) \cdot A_{mol,1}(0) + A^2_{Au}(0) - A^2_{L,N}(0) = 0 \tag{7.6}$$

Using the Guinier extrapolated $I_{Au}(0)$ as $A^2_{Au}(0)$ and $I_{L,N}(0)$ as $A^2_{L,N}(0)$ this quadratic equation can be solved for $A_{mol}(0)$; the solution with the negative root for $A_{mol}(0)$ value is neglected since negative values for $A_{mol}(0)$ are non-physical in this context. Using the computed value A_{mol} from Eq. 7.6, the scaling value cunlabeld was calculated as the ratio between the computed A^2_{mol} and the Guinier extrapolated $I_{mol}(0)$ of the experimental data for the unlabeled sample to scale the experimental profile of the unlabeled sample for the further analysis described below.

$$c_{unlabeled} = \frac{A^2_{mol,1}(0)}{I_{mol}(0)} \tag{7.7}$$

$$I(s)_{mol,N} = c_{unlabeled} \cdot I_{mol}(s) \tag{7.8}$$

7.4.9 Determination of the gold label position in the low-resolution structure

Our method requires a set of trial gold label positions enveloping the low-resolution electron density map and compares the scattering profile for placing the gold label at each individual trial point to the experimental data for the gold-labeled macromolecule. In order to obtain a set of possible gold label positions, 100 random coordinates on a sphere of 11 Å radius (corresponding to the 7 Å gold nanocrystal radius and the approximately 4 Å linker length [69, 82, 105, 139] were generated around each dummy atom or bead in the reconstruction. Next, sterically forbidden positions were eliminated from this set of trial gold positions, by discarding all trial positions that were located within 7 Å (the nanocrystal size) to any bead of the reconstructed model. Control calculations with fewer or more trial positions per bead (Fig. 7.14) suggest that 100 trial positions provide sufficient accuracy for the assignment of best label positions, while being computationally efficient.

Next, theoretical scattering profiles for the labeled molecule with the gold label in each of the trial gold positions are calculated. This computation can be carried out efficiently, since the scattering cross-term caused by interference between the gold nanocrystal and the macromolecule is the only term affected by the label position relative to the molecule. The intensity of the cross term is calculated using the Debye formula (Eq. 7.1), including only gold-molecules terms, i.e., the index i runs over all beads or point scatterers in the representation of the macromolecule (see Equation 7.2) and r_i is the distance from bead i to the center of the gold label:

$$I_{Au-mol}(s) = \sum_{i=1}^{N} 2 \cdot A_{Au}(s) \cdot A_{mol} \frac{\sin(2 \cdot \pi \cdot s \cdot r_i)}{2 \cdot \pi \cdot s \cdot r_i} \tag{7.9}$$

The scattering amplitude $A_{Au}(s)$ is determined by taking the square root of the non-negative least square fit to the experimental measure data of the gold nanocrystals.

$$A_{Au}(s) = \sqrt{I_{Au}(s)} \tag{7.10}$$

This procedure takes into account the angle dependency of the gold particle scattering. We note that Equation 7.10 neglects the imaginary part of the atomic form factors, i.e., of the anomalous dispersion. However, contributions from the imaginary part of the dispersion are negligible for our measurements, since we are sufficiently far (≥80 eV) from the absorption edges of all atomic species involved. Moreover, the molecular scattering amplitude A_{mol}, approximated as s-independent, is calculated via the following expression.

$$A_{mol} = \left(\frac{c_1 \cdot I_{mol}^{model}(s_1) + c_2}{N^2} \right)^{1/2} \tag{7.11}$$

The scaling values c_1 and c_2 are obtained from fitting the Debye formula (Equation 7.2). N is given by the number of dummy atoms in the shape reconstruction and s_1 by the smallest scattering angle s appearing in the interpolated experimental data.

After computing the cross-scattering term I_{Au-mol} for every trial gold label position using Eqs. 7.9 to 7.11, the cross-scattering intensity profiles are summed with the reconstructed gold

label I_{Au} and unlabeled molecule I_{mol} scattering term in order to get a reconstructed scattering profile of the labeled sample I_{total} (Fig. 7.1b).

$$I_{total} = I_{Au}(s) + I_{mol}(s) + I_{Au-mol}(s) \tag{7.12}$$

To determine the most likely label position, we compare the computed scattering profiles for the labeled molecule I_{total} to the experimentally measured scattering profiles of the labeled samples using a following criterion:

$$T = \frac{\sum_s \left[c_{min,1} \cdot I_{total}(s) + c_{min,2} - I_{L,N}(s)\right]^2}{\sum_s I_{L,N}^2(s)} \tag{7.13}$$

where $I_{L,N}$ is the experimental measured (normalized) scattering profile of the gold labeled sample and I_{Au}, I_{mol}, and I_{Au-mol} are the reconstructed scattering intensity profiles of the gold nanocrystals, the unlabeled macromolecule, and the cross-scattering term, respectively. The scaling parameters $c_{min,1}$ and $c_{min,2}$ were determined for every trial position using the fminsearch function in MATLAB to minimize T. Finally, all trial label positions were sorted by their T scores from best to worst fitting.

7.4.10 Filtering of gold label positions

For some of the samples, the best fitting gold-label positions were further processed using the following clustering approach. First, we compute the mean (X,Y,Z)-position for the 10 best fitting trial positions. Next, we evaluate the distance from the mean (X,Y,Z)-position and the standard deviation in each Carthesian coordinate for the 100 best fitting trial positions. Points are eliminated if their distance from the mean (X,Y,Z)-position is greater than one standard deviation in any of the three coordinates. The remaining points are used for further analysis.

7.5 Supplementary Materials

7.5.1 Supplementary Figures

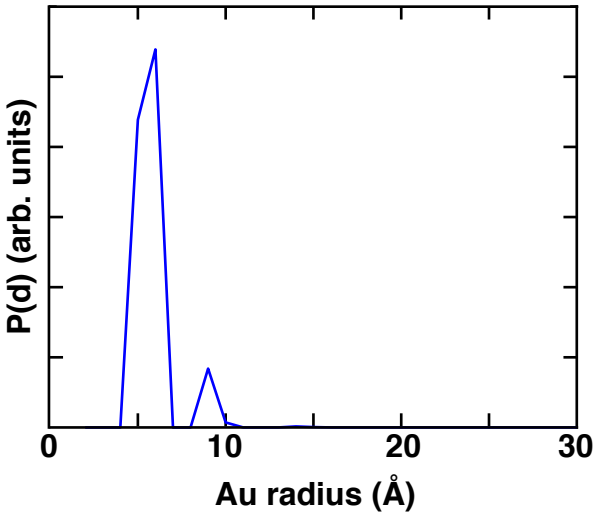

Figure 7.7: Radius distribution of gold nanocrystals. (Volume-weighted) radius distribution of a typical gold nanocrystal batch determined from SAXS (see Methods 7.4). The nanoparticles are monodisperse with a radius centered at 6-7 Å.

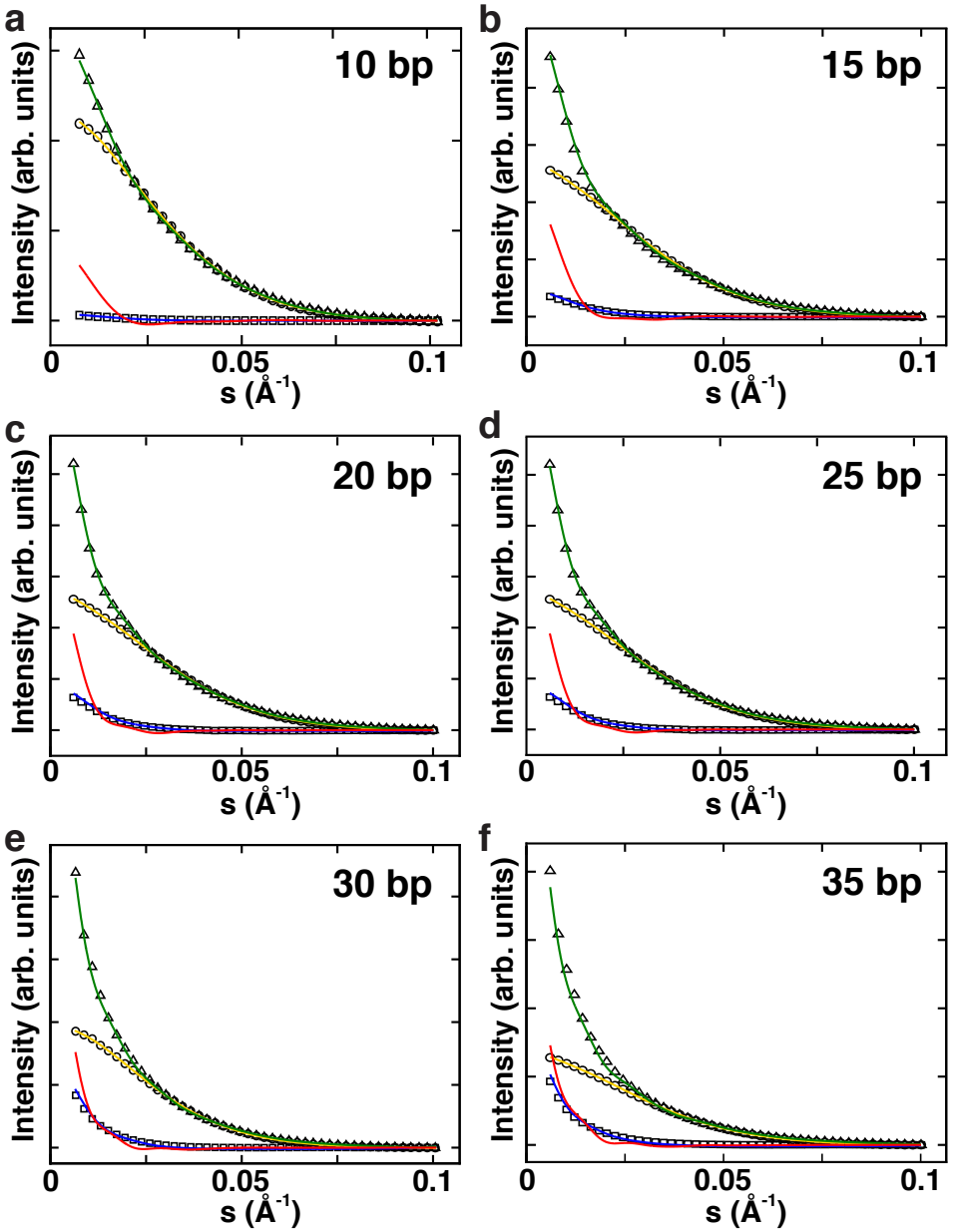

Figure 7.8: SAXS scattering profiles of end-labeled DNA samples for label position reconstruction. The panels show experimental and reconstructed data for (**a**) 10 bp, (**b**) 15 bp, (**c**) 20 bp, (**d**) 25 bp, (**e**) 30 bp and (**f**) 35 bp DNA. Experimentally measured profiles of single labeled molecules (black triangles), gold nanocrystals in solution (black circles), and unlabeled DNA molecules (black squares) are displayed in each panel. The number of s-bins shown was reduced for clarity. Reconstructed profiles of single labeled DNA (green line), gold nanocrystals in solution (yellow line), unlabeled DNA (blue line) and the cross-scattering term between gold nanocrystal and molecule (red line) are shown in each panel. The intensity ratio at low s-values between unlabeled molecule and gold nanocrystal is clearly decreasing with higher molecular mass of the molecule (from (a) to (f)).

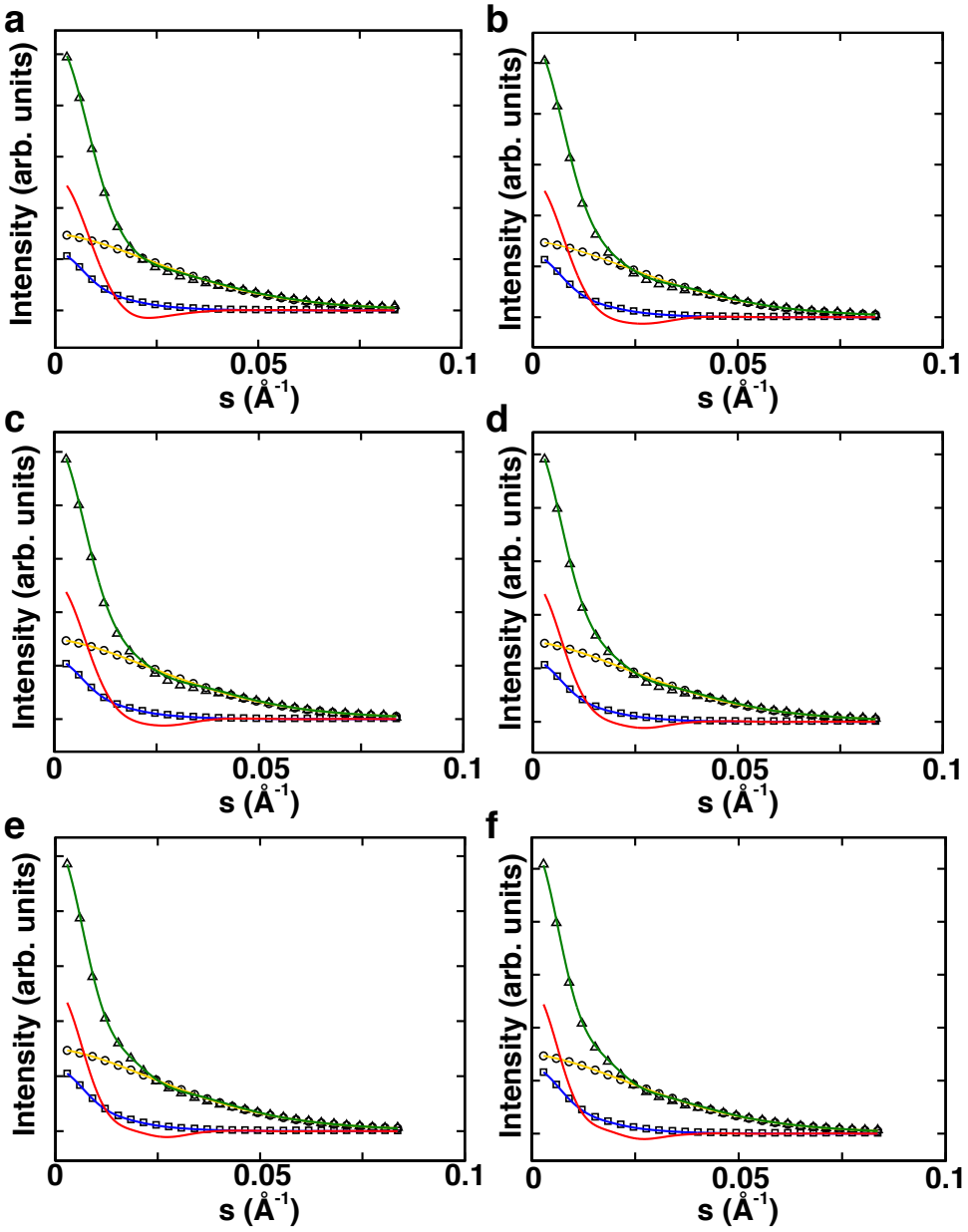

Figure 7.9: SAXS scattering profiles of the full set of DNA samples for internal label position reconstruction. The panels show experimental and reconstructed data for (**a**) B11, (**b**) B17, (**c**) A8, (**d**) A19, (**e**) A21 and (**f**) A4 DNA (see main text 7.2.2 and Table 7.2). Experimentally measured profiles of single labeled molecules (black triangles), gold nanocrystals in solution (black circles), and unlabeled DNA molecules (black squares) are displayed in each panel. The number of *s*-bins shown was reduced for clarity. Reconstructed profiles of single labeled DNA (green line), gold nanocrystals in solution (yellow line), unlabeled DNA (blue line) and the cross-scattering term between gold nanocrystal and molecule (red line) are shown in each panel.

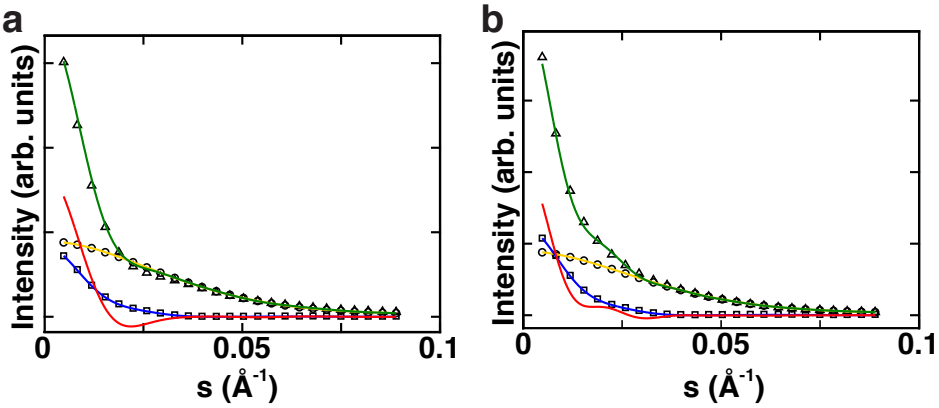

Figure 7.10: SAXS scattering profiles of two end labeled RNA kink-turn motif samples. The shown panels contain experimental and reconstructed data for (**a**) A-strand and (**b**) B-strand labeled constructs. Experimentally measured profiles of single labeled molecules (black triangles), gold nanocrystals in solution (black circles), and unlabeled DNA molecules (black squares) are displayed in each panel. The number of *s*-bins shown was reduced for clarity. Reconstructed profiles of single labeled RNA (green line), gold nanocrystals in solution (yellow line), unlabeled RNA (blue line) and the cross-scattering term between gold nanocrystal and molecule (red line) are shown in each panel.

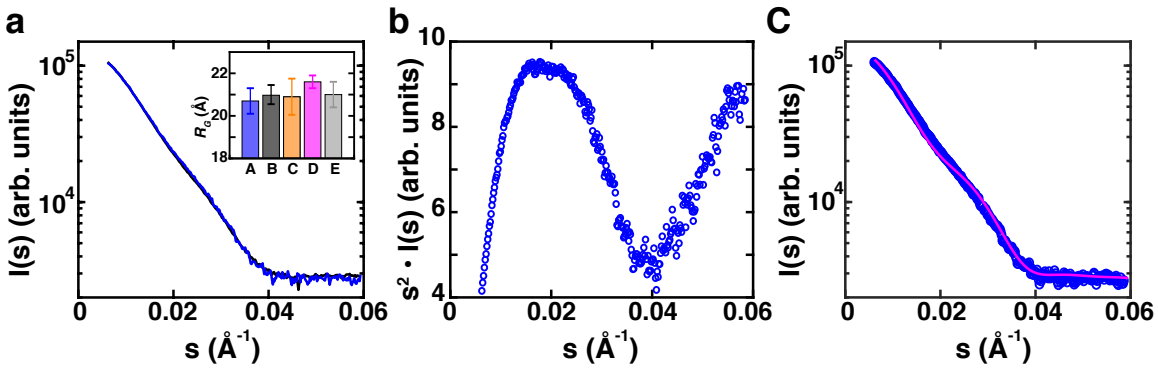

Figure 7.11: SAXS data for unlabeled calmodulin. (**a**) Comparison of experimental scatter profiles for calmodulin recorded in the presence of 5 mM $CaCl_2$ (blue) or 10 mM $CaCl_2$ (black). The inset shows the radii of gyration R_G from different sets of experimental data in the presence of Ca^{2+}. Data sets are: (A): This work, 5 mM $CaCl_2$ (blue); (B): This work, 10 mM $CaCl_2$ (black); (C): Seaton *et al.* [184] (orange); (D): Kataoka *et al.* [191] (magenta); (E): Heidorn and Trewhella [190] (grey). R_G values and errors were calculated in this work via Guinier analysis using Primus [60]. Errors for literature values were estimated calculating the standard deviation between values given for Guinier analysis and extrapolation to zero concentration (orange, grey) or Guiner analysis and indirect Fourier transformation (magenta). (**b**) Kratky plot of calmodulin scattering data recorded in the presence of 5 mM $CaCl_2$. (**c**) Experimental scattering data of calmodulin (blue circles) and scattering profile predicted from a crystal structure of a homologous (~90 % sequence identity) calmodulin using FoXS [193] (magenta line; PDB code: 1EXR [73]).

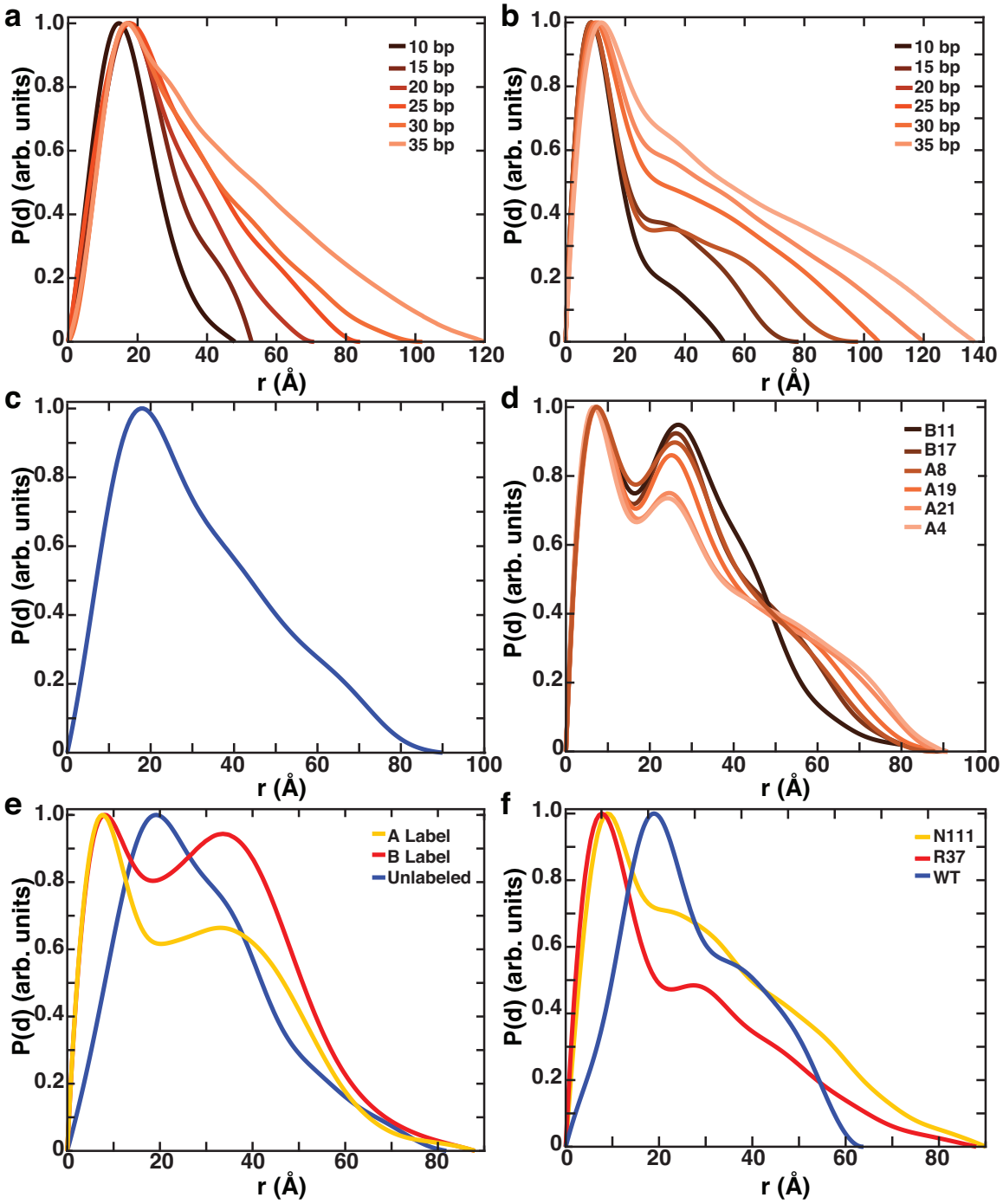

Figure 7.12: Pair-distance distributions for all unlabeled and labeled samples. (**a**) Pair-distance distributions for unlabeled DNA from 10 bp to 35 bp. (**b**) Pair-distance distributions for end-labeled DNA from 10 bp to 35 bp. (**c**) Pair-distance distribution for unlabeled 26 bp DNA. (**d**) Pair-distance distributions for internal labeled 26 bp DNA. (**e**) Pair-distance distributions for the unlabeled RNA kink-turn (blue), the labeled A strand (dark yellow) and the labeled B strand (red). (**f**) Pair-distance distributions for the unlabeled wildtype of calmodulin (blue), the N111C (dark yellow) labeled and R37C (red) labeled construct. The pair-distance distributions were calculated using Primus [60] and normalized by dividing by the maximum value in the distribution.

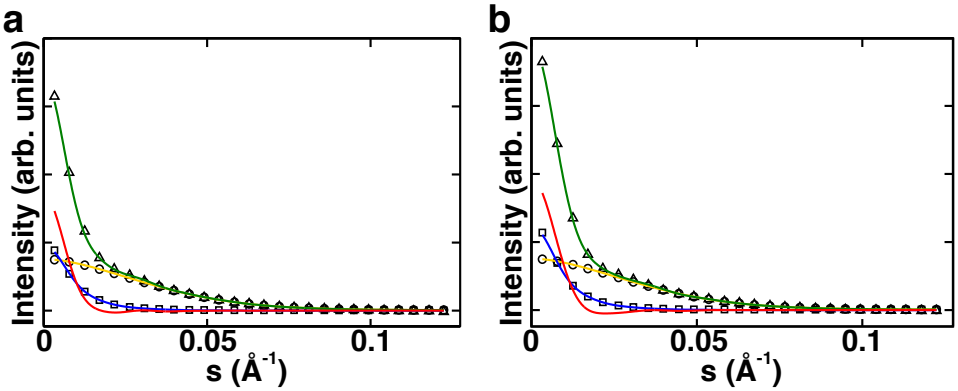

Figure 7.13: SAXS scattering profiles of two different labeled calmodulin samples. The panels show experimental and reconstructed data for (**a**) N111C and (**a**) R37C. Experimentally measured profiles of single labeled molecules (black triangles), gold nanocrystals in solution (black circles), and unlabeled DNA molecules (black squares) are displayed in each panel. Measured data sets were reduced for clarity. Reconstructed profiles of single labeled DNA (green line), gold nanocrystals in solution (yellow line), unlabeled DNA (blue line) and the cross-scattering term between gold nanocrystal and molecule (red line) are shown in each panel.

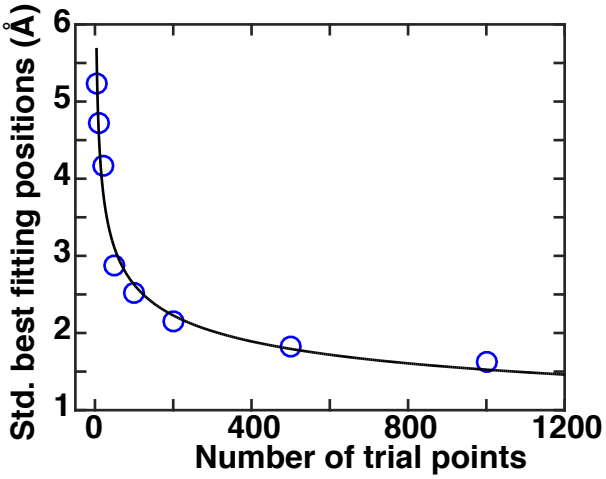

Figure 7.14: Spread of the best fitting position depending on the initial trial point density. Sets from 5 to 1000 initial trial points per bead in the reconstructed molecular density map were created for one exemplary internally-labeled 26 bp DNA construct. The best fitting positions were calculated and the standard deviation along the helical axis (see Results 7.2) was calculated for the best 100 fitting positions. The black line is a fit of a one-component exponential equation..

7.5.2 Supplementary Tables

DNA construct	Sequence
10 bp	5'–GCATCTGGGC–3' CGTAGACCCG
15 bp	5'–CGACTCTACGGAAGG–3' GCTGAGATGCCTTCC
20 bp	5'–CGACTCTACGGCATCTGCGC–3' GCTGAGATGCCGTAGACGCG
25 bp	5'–CGACTCTACGGAAGGGCATCTGCGC–3' GCTGAGATGCCTTCCCGTAGACGCG
30 bp	5'–CGACTCTACGGAAGGTCTCGGACTACGCGC–3' GCTGAGATGCCTTCCAGAGCCTGATGCGCG
35 bp	5'–CGACTCTACGGAAGGGCATCTCTCGGACTACGCGC–3' GCTGAGATGCCTTCCCGTAGAGAGCCTGATGCGCG

Table 7.1: DNA sequences used in the experiments on end-labeled DNA.

DNA construct	Sequence
26 bp A4 – internal	5' – CGATCCGTGAAGGCGATCTCTGCGGC – 3' GCTAGGCACTTCCGCTAGAGACGCCG
26 bp A8 – internal	5' – CGATCCGTGAAGGCGATCTCTGCGGC – 3' GCTAGGCACTTCCGCTAGAGACGCCG
26 bp A19 – internal	5' – CGATCCGTGAAGGCGATCTCTGCGGC – 3' GCTAGGCACTTCCGCTAGAGACGCCG
26 bp A21 – internal	5' – CGATCCGTGAAGGCGATCTCTGCGGC – 3' GCTAGGCACTTCCGCTAGAGACGCCG
26 bp B11 – internal	5' – CGATCCGTGAAGGCGATCTCTGCGGC – 3' GCTAGGCACTTCCGCTAGAGACGCCG
26 bp B17 – internal	5' – CGATCCGTGAAGGCGATCTCTGCGGC – 3' GCTAGGCACTTCCGCTAGAGACGCCG

Table 7.2: DNA sequences used in the experiments with internally-labeled DNA. The bases labeled in red font denote the modified position for the gold nanocrystal attachment. Sequences for strands A are shown on top, for strand B on bottom.

RNA construct	Sequence
27 bp	5' – GCUCUCUCCAGCGAAGAACUUGGUUGC – 3' CGAGAGAGGUCG - - - AGGGAACCAACG

Table 7.3: RNA kink-turn sequence. Sequences for strands A are shown on top, for strand B on bottom.

Sample:	Radius of gyration R_g (Å):
10 bp unlabeled	14.0 ± 0.2
10 bp labeled	11.7 ± 0.1
15 bp unlabeled	17.6 ± 0.3
15 bp labeled	17.5 ± 0.3
20 bp unlabeled	19.8 ± 0.4
20 bp labeled	21.1 ± 0.5
25 bp unlabeled	21.6 ± 1.3
25 bp labeled	18.5 ± 0.6
30 bp unlabeled	24.3 ± 2.4
30 bp labeled	24.2 ± 2.4
35 bp unlabeled	28.6 ± 2.3
35 bp labeled	26.9 ± 1.4
26 bp unlabeled	24.7 ± 0.4
B11 internal label	21.0 ± 0.3
B17 internal label	22.0 ± 0.3
A8 internal label	21.9 ± 0.3
A19 internal label	22.8 ± 0.3
A21 internal label	23.7 ± 0.4
A4 internal label	24.2 ± 0.4
RNA kink-turn unlabeled	27.9 ± 0.7
RNA kink-turn A label	24.2 ± 0.5
RNA kink-turn B label	25.6 ± 0.6
Calmodulin unlabeled	20.7 ± 0.4
Calmodulin R37C label	19.8 ± 0.8
Calmodulin N111C label	23.9 ± 0.6

Table 7.4: Radius of gyration R_g values for all unlabeled and labeled samples. The radius of gyration values were calculated from Guinier plots using Primus [60].

Part III

Conclusion and Outlook

Chapter 8
Conclusion and Outlook

The function of biological macromolecules and the complexes they form is governed by the structural solution ensemble they adopt under certain solution and environmental conditions. A better understanding of the underlying folding pathways is required to reconstruct the free-energy landscape and the impact on vast parts of basic biology such as allosteric regulation, signaling or translation. To study macromolecular structures and the conformational ensembles they adopt small-angle X-ray scattering (SAXS) and the emerging technique X-ray scattering interferometry (XSI) has been established with unique advantages and capabilities.

Within this thesis I have contributed to further develop the application of gold labeled macromolecules to enhance the information content gained by small-angle X-ray scattering experiments.

The first part of this thesis covers the design and preparation of end-labeled nucleic acid samples. It consists of three detailed protocols that are required to prepare samples, to acquire XSI data, and to generate ensemble distance distributions. While the first protocol describes sample preparation for end-labeled nucleic acid gold conjugates and includes a protocol for gold nanocrystal synthesis, the second part focuses on the acquisition of a full data set at a synchrotron radiation facility for XSI analysis. Finally, the third protocol describes the data analysis of XSI data and the use of a custom-written graphical user interface (GUI) in MATLAB. The detailed protocols and the user interface presented in this thesis will enable scientists interested in molecular distance measurements to perform and analyze XSI measurements easily.

Further, in this thesis I have shown that a novel anomalous small-angle X-ray scattering interferometry method has the advantage that only the double labeled sample needs to be prepared and measured and that it does not rely on intrinsic assumptions about the macromolecular scattering contributions as compared to previously reported X-ray scattering interferometry using multiple samples. These properties will immediately allow for measuring macromolecular samples without the need for orthogonal labeling strategies. While these advantages make the approach reliable and experimentally attractive, yet still absolute distances with Ångström resolution and precise distance distributions to evaluate the flexibility of macromolecular systems can be obtained. In the course of this thesis gold-labeled DNA samples were used, however it should be possible to readily expand the method to labeled proteins or similar to conventional XSI to protein-nucleic acid complexes. Moreover, other heavy metals can serve as nanoparticle

component which allows for orthogonal labeling options and might permit to measure several distinct distance distributions for one sample. Combined, ASAXS and gold nanoparticles provide a powerful new approach to determining intra-molecular distance distributions for labeled biological macromolecules, an approach that is anticipated to provide new and quantitative insights into the structure, dynamics, and interactions of biological macromolecules.

Furthermore, the application of XSI measurements to investigate the solution ensemble of the Holliday junction is presented. The XSI results agree well with prior FRET and gel-mobility studies for high salt, showing that the four-way junction preferentially adopts two distinct X-structures with coaxial pairwise stacking of helical arms. Thus, XSI experiments performed under a range of solution conditions suggest that the ionic strength of the buffer environment as well as the bases located in the center of the structure can dramatically alter the free-energy landscape of the Holliday junction.

The final chapter of this thesis focused on SAXS measurements to determine gold nanocrystal label positions relative to low-resolution electron density maps computed from SAXS data. The findings include 10-35 bp long end-labeled dsDNA, internally labeled dsDNA, an RNA kink-turn motif, and calmodulin as a protein model system. The results show that this novel method is broadly applicable to all major classes of biological macromolecules. The approach combines low-resolution electron density maps computed from SAXS data and small gold nanocrystals as fiducial markers. Determining the 3D position of specific labeled residues allows the sequence of the macromolecule to be placed within the macromolecular shape and to distinguish different domains as well as secondary structure elements. In summary, this method is a powerful new approach to determining precise residue positions in biological macromolecules.

The work presented in this thesis opens up several exciting possibilities such as to extend X-ray scattering interferometry gold-gold distance measurements by SAXS or ASAXS to proteins, especially for folding intermediates and intrinsically disordered proteins (IDPs). Further, there are new opportunities in the analysis of the structural arrangement of protein-nucleic acid complexes and probing single-stranded nucleic acid molecules. Moreover, the dependence of these molecules on environmental conditions, such as varying ionic strength or the binding of ligands, can be probed. Combining the reconstruction method and X-ray scattering interferometry could provide additional information on reconstructions that have, for example, an internal symmetry, in particular to resolve ambiguities in the relative placements of residues. Furthermore, these methods are not limited to gold nanocrystals labels but the techniques can be expect to be equally applicable to other labels, such as silver and platinum nanoparticles. While this proof-of-concept study combined single labeled samples and small-angle X-ray scattering site specifically labeling of proteins using gold nanoparticles might provide an attractive approach for multiple techniques as for example cryo-EM. Finally, the application of next-generation free electron lasers with very high brilliance could allow measurements of correlated and time-resolved distances between multiple sites in a macromolecule with a time resolution of tens of femtoseconds.

Part IV

Appendix

Bibliography

[1] T. Zettl, R. Das, P. A. B. Harbury, D. Herschlag, J. Lipfert, R. S. Mathew, and X. Shi, Recording and Analyzing Nucleic Acid Distance Distributions with X-ray Scattering Interferometry (XSI). *Current Protocols in Nucleic Acid Chemistry* (in Press).

[2] T. Zettl, R. S. Mathew, S. Seifert, S. Doniach, P. A. Harbury, and J. Lipfert, Absolute Intra-Molecular Distance Measurements with Ångström-Resolution using Anomalous Small-Angle X-ray Scattering. *Nano Letters*, **16**(9):5353–5357 (2016), doi:10.1021/acs.nanolett.6b01160.

[3] T. Zettl, R. S. Mathew, X. Shi, S. Doniach, D. Herschlag, P. A. B. Harbury, and J. Lipfert, Gold Nanocrystal Labels Provide a Sequence-to-3D Structure Map in SAXS Reconstructions. *Science Advances*, **4**:eaar4418 (2018), doi:10.1126/sciadv.aar4418.

[4] C. Huang and C. G. Kalodimos, Structures of Large Protein Complexes Determined by Nuclear Magnetic Resonance Spectroscopy. *Annual Review of Biophysics*, **46**(1):317–336 (2017), doi:10.1146/annurev-biophys-070816-033701.

[5] S. Subramaniam, L. A. Earl, V. Falconieri, J. L. S. Milne, and E. H. Egelman, Resolution advances in cryo-EM enable application to drug discovery. *Current Opinion in Structural Biology*, **41**(Supplement C):194–202 (2016), doi:https://doi.org/10.1016/j.sbi.2016.07.009.

[6] H.-W. Wang and J.-W. Wang, How cryo-electron microscopy and X-ray crystallography complement each other. *Protein Science*, **26**(1):32–39 (2017), doi:10.1002/pro.3022.

[7] H. N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz, D. P. DePonte, U. Weierstall, R. B. Doak, F. R. N. C. Maia, A. V. Martin, I. Schlichting, L. Lomb, N. Coppola, R. L. Shoeman, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, L. Foucar, N. Kimmel, G. Weidenspointner, P. Holl, M. Liang, M. Barthelmess, C. Caleman, S. Boutet, M. J. Bogan, J. Krzywinski, C. Bostedt, S. Bajt, L. Gumprecht, B. Rudek, B. Erk, C. Schmidt, A. Hömke, C. Reich, D. Pietschner, L. Strüder, G. Hauser, H. Gorke, J. Ullrich, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K.-U. Kühnel, M. Messerschmidt, J. D. Bozek, S. P. Hau-Riege, M. Frank, C. Y. Hampton, R. G. Sierra, D. Starodub, G. J. Williams, J. Hajdu, N. Timneanu, M. M. Seibert, J. Andreasson, A. Rocker, O. Jönsson, M. Svenda, S. Stern, K. Nass, R. Andritschke, C.-D. Schröter, F. Krasniqi, M. Bott, K. E. Schmidt, X. Wang, I. Grotjohann, J. M. Holton, T. R. M. Barends, R. Neutze, S. Marchesini, R. Fromme, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, I. Andersson, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson, and J. C. H. Spence, Femtosecond X-ray protein nanocrystallography. *Nature*, **470**:73 (2011), doi:10.1038/nature09750https://www.nature.com/articles/nature09750#supplementary-information.

[8] S. Boutet, L. Lomb, G. J. Williams, T. R. M. Barends, A. Aquila, R. B. Doak, U. Weierstall, D. P. DePonte, J. Steinbrener, R. L. Shoeman, M. Messerschmidt, A. Barty, T. A. White,

S. Kassemeyer, R. A. Kirian, M. M. Seibert, P. A. Montanez, C. Kenney, R. Herbst, P. Hart, J. Pines, G. Haller, S. M. Gruner, H. T. Philipp, M. W. Tate, M. Hromalik, L. J. Koerner, N. van Bakel, J. Morse, W. Ghonsalves, D. Arnlund, M. J. Bogan, C. Caleman, R. Fromme, C. Y. Hampton, M. S. Hunter, L. C. Johansson, G. Katona, C. Kupitz, M. Liang, A. V. Martin, K. Nass, L. Redecke, F. Stellato, N. Timneanu, D. Wang, N. A. Zatsepin, D. Schafer, J. Defever, R. Neutze, P. Fromme, J. C. H. Spence, H. N. Chapman, and I. Schlichting, High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. *Science*, **337**:362–364 (2012).

[9] G. Zanotti, Cryo-EM and X-Ray Crystallography: Complementary or Alternative Techniques? *NanoWorld Journal*, **2**(2):22–23 (2016), doi:10.17756/nwj.2016-025.

[10] M. Pfreundschuh, D. Martinez-Martin, E. Mulvihill, S. Wegmann, and D. J. Muller, Multiparametric high-resolution imaging of native proteins by force-distance curve–based AFM. *Nature Protocols*, **9**:1113 (2014), doi:10.1038/nprot.2014.070.

[11] D. J. Müller, F. A. Schabert, G. Büldt, and A. Engel, Imaging purple membranes in aqueous solutions at sub-nanometer resolution by atomic force microscopy. *Biophysical Journal*, **68**(5):1681–1686 (1995), doi:https://doi.org/10.1016/S0006-3495(95)80345-0.

[12] S. Karrasch, R. Hegerl, J. H. Hoh, W. Baumeister, and A. Engel, Atomic force microscopy produces faithful high-resolution images of protein surfaces in an aqueous environment. *Proceedings of the National Academy of Sciences of the United States of America*, **91**(3):836–838 (1994), doi:10.1073/pnas.91.3.836.

[13] C. D. Putnam, M. Hammel, G. L. Hura, and J. A. Tainer, X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics*, **40**(3):191–285 (2007), doi:10.1017/s0033583507004635.

[14] P. Chacón, F. Morán, J. F. Díaz, E. Pantos, and J. M. Andreu, Low-Resolution Structures of Proteins in Solution Retrieved from X-Ray Scattering with a Genetic Algorithm. *Biophysical Journal*, **74**(6):2760–2775 (1998), doi:https://doi.org/10.1016/S0006-3495(98)77984-6.

[15] D. I. Svergun, Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical Journal*, **76**(6):2879–2886 (1999), doi:S0006-3495(99)77443-6[pii]10.1016/S0006-3495(99)77443-6.

[16] D. Walther, F. E. Cohen, and S. Doniach, Reconstruction of low-resolution three-dimensional density maps from one-dimensional small-angle X-ray solution scattering data for biomolecules. *Journal of Applied Crystallography*, **33**:350–363 (2000).

[17] A. Guinier, La diffraction des rayons X aux très petits angles: Application à l'étude de phénomènes ultramicroscopiques. *Annals of Physics (Paris)*, **12**:161–237 (1939).

[18] M. V. Petoukhov and D. I. Svergun, Applications of small-angle X-ray scattering to biomacromolecular solutions. *The International Journal of Biochemistry & Cell Biology*, **45**(2):429–437 (2013), doi:https://doi.org/10.1016/j.biocel.2012.10.017.

[19] J. Lipfert, I. S. Millett, S. Seifert, and S. Doniach, Sample holder for small-angle x-ray scattering static and flow cell measurements. *Review of Scientific Instruments*, **77**(4):26–28 (2006).

[20] L. K. Brützel, T. Gerling, S. M. Sedlak, P. U. Walker, W. Zheng, H. Dietz, and J. Lipfert, Conformational Changes and Flexibility of DNA Devices Observed by Small-Angle X-ray Scattering. *Nano Letters*, **16**(8):4871–9 (2016), doi:10.1021/acs.nanolett.6b01338.

[21] D. I. Svergun, M. V. Petoukhov, and M. H. J. Koch, Determination of Domain Structure of Proteins from X-Ray Solution Scattering. *Biophysical Journal*, **80**(6):2946–2953 (2001), doi:https://doi.org/10.1016/S0006-3495(01)76260-1.

[22] J. Lipfert, V. B. Chu, Y. Bai, D. Herschlag, and S. Doniach, Low-resolution models for nucleic acids from small-angle X-ray scattering with applications to electrostatic modeling. *Journal of Applied Crystallography*, **40**:229–234 (2007).

[23] D. Franke and D. I. Svergun, DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography*, **42**(2):342–346 (2009), doi:10.1107/S0021889809000338.

[24] M. V. Petoukhov and D. I. Svergun, Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophysical Journal*, **89**(2):1237–50 (2005), doi:S0006-3495(05)72771-5[pii]10.1529/biophysj.105.064154.

[25] H. T. Nguyen, S. A. Pabit, S. P. Meisburger, L. Pollack, and D. A. Case, Accurate small and wide angle x-ray scattering profiles from atomic models of proteins and nucleic acids. *The Journal of chemical physics*, **141**(22):22D508 (2014), doi:10.1063/1.4896220.

[26] H. T. Nguyen, S. A. Pabit, L. Pollack, and D. A. Case, Extracting water and ion distributions from solution x-ray scattering experiments. *The Journal of Chemical Physics*, **144**(21):214105 (2016), doi:10.1063/1.4953037.

[27] C. E. M. Schindler, S. J. de Vries, A. Sasse, and M. Zacharias, SAXS Data Alone can Generate High-Quality Models of Protein-Protein Complexes. *Structure*, **24**(8):1387–1397 (2016), doi:https://doi.org/10.1016/j.str.2016.06.007.

[28] P. Willmott, *An Introduction to Synchrotron Radiation: Techniques and Applications*, book chapter The Interaction of X-rays with Matter, pages 15–37, John Wiley & Sons, Ltd, Chichester, UK (2011), ISBN 9781119970958, doi:10.1002/9781119970958.ch2.

[29] N. Stribeck, *X-Ray Scattering of Soft Matter*, book chapter General Background, pages 7–32, Springer Berlin Heidelberg, Berlin, Heidelberg (2007), ISBN 978-3-540-69856-2, doi:10.1007/978-3-540-69856-2_2.

[30] J. Als-Nielsen and D. McMorrow, *Elements of Modern X-ray Physics*, book chapter X-rays and their interaction with matter, pages 1–28, John Wiley & Sons, Inc., Hoboken, NJ, USA (2011), ISBN 9781119998365, doi:10.1002/9781119998365.ch1.

[31] D. I. Svergun, M. H. J. Koch, P. A. Timmins, and R. P. May, *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*, Oxford University Press (2013), doi:10.1093/acprof:oso/9780199639533.001.0001.

[32] P. Willmott, *An Introduction to Synchrotron Radiation: Techniques and Applications*, book chapter Scattering Techniques, pages 133–221, John Wiley & Sons, Ltd, Chichester, UK (2011), ISBN 9781119970958, doi:10.1002/9781119970958.ch5.

[33] J. Als-Nielsen and D. McMorrow, *Elements of Modern X-ray Physics*, book chapter Kinematical scattering I: non-crystalline materials, pages 113–146, John Wiley & Sons, Inc., Hoboken, NJ, USA (2011), ISBN 9781119998365, doi:10.1002/9781119998365.ch4.

[34] L. K. Bruetzel, S. Fischer, A. Salditt, S. M. Sedlak, B. Nickel, and J. Lipfert, A Mo-anode-based in-house source for small-angle X-ray scattering measurements of biological macromolecules. *Review of Scientific Instruments*, **87**(2):025103 (2016).

[35] F. de Bergevin, *X-ray and Neutron Reflectivity: Principles and Applications*, book chapter The Interaction of X-Rays (and Neutrons) with Matter, pages 1–57, Springer Berlin Heidelberg, Berlin, Heidelberg (2009), ISBN 978-3-540-88588-7, doi:10.1007/978-3-540-88588-7_1.

[36] A. Guinier and G. Fournet, *Small-angle scattering of X-rays*, John Wiley & Sons, Inc., New York, NY (1955).

[37] P. Debye, Zerstreuung von Röntgenstrahlen. *Annalen der Physik*, **351**(6):809–823 (1915), doi:10.1002/andp.19153510606.

[38] H. B. Stuhrmann, Ein neues Verfahren zur Bestimmung der Oberflächenform und der inneren Struktur von gelösten globulären Proteinen aus Röntgenkleinwinkelmessungen. *Zeitschrift für Physikalische Chemie*, **72**(4_6):177–184 (1970), doi:10.1524/zpch.1970.72.4_6.177.

[39] H. Stuhrmann, Interpretation of small-angle scattering functions of dilute solutions and gases. A representation of the structures related to a one-particle scattering function. *Acta Crystallographica Section A*, **26**(3):297–306 (1970), doi:10.1107/S0567739470000748.

[40] S. Doniach, Changes in biomolecular conformation seen by small angle X-ray scattering. *Chemical Reviews*, **101**(6):1763–78 (2001), doi:cr990071k[pii].

[41] G. L. Hura, C. L. Tsai, S. A. Claridge, M. L. Mendillo, J. M. Smith, G. J. Williams, A. J. Mastroianni, A. P. Alivisatos, C. D. Putnam, R. D. Kolodner, and J. A. Tainer, DNA conformations in mismatch repair probed in solution by X-ray scattering from gold nanocrystals. *Proceedings of the National Academy of Sciences of the United States of America*, **110**(43):17308–13 (2013), doi:10.1073/pnas.1308595110.

[42] A. Grishaev, M. G. Clore, and N. J. Anthis, Contrast-Matched Small-Angle X-ray Scattering from a Heavy-Atom- Labeled Protein in Structure Determination: Application to a Lead-Substituted Calmodulin–Peptide Complex. *Journal of the American Chemical Society*, pages 14686–14689 (2012).

[43] J. Lipfert and S. Doniach, Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annual Review of Biophysics and Biomolecular Structure*, **36**:307–327 (2007), doi:10.1146/annurev.biophys.36.040306.132655.

[44] J. Als-Nielsen and D. McMorrow, *Elements of Modern X-ray Physics*, book chapter Photoelectric absorption, pages 239–273, John Wiley & Sons, Inc., Hoboken, NJ, USA (2011), ISBN 9781119998365, doi:10.1002/9781119998365.ch7.

[45] National Institute of Standards and Technology (NIST), Mass energy-absorption coefficient of gold (Online; accessed: 03-01-2018), URL `https://physics.nist.gov/PhysRefData/XrayMassCoef/ElemTab/z79.html`.

[46] D. Eichert, *Synchrotron Radiation: Basics, Methods and Applications*, book chapter X-Ray Microscopy, pages 409–436, Springer Berlin Heidelberg, Berlin, Heidelberg (2015), ISBN 978-3-642-55315-8, doi:10.1007/978-3-642-55315-8_14.

[47] P. Fornasini, *Synchrotron Radiation: Basics, Methods and Applications*, book chapter Introduction to X-Ray Absorption Spectroscopy, pages 181–211, Springer Berlin Heidelberg, Berlin, Heidelberg (2015), ISBN 978-3-642-55315-8, doi:10.1007/978-3-642-55315-8_6.

[48] H. B. Stuhrmann and H. Notbohm, Configuration of the four iron atoms in dissolved human hemoglobin as studied by anomalous dispersion. *Proceedings of the National Academy of Sciences of the United States of America*, **78**(10):6216–20 (1981).

[49] J. Als-Nielsen and D. McMorrow, *Elements of Modern X-ray Physics*, book chapter Resonant scattering, pages 275–303, John Wiley & Sons, Inc., Hoboken, NJ, USA (2011), ISBN 9781119998365, doi:10.1002/9781119998365.ch8.

[50] D. Creagh, Tables of X-ray absorption corrections and dispersion corrections : the new versus the old. *Chemical Physics*, **C**:417–434 (1990), doi:10.1016/0168-9002(90)90722-I.

[51] Biomolecular Structure Center at University of Washington, X-ray Absorption Edges (Online; accessed 08-01-2018), URL `http://skuld.bmsc.washington.edu/scatter/AS_periodic.html`.

[52] D. Creagh, Theoretical and experimental techniques for the determination of X-ray anomalous dispersion corrections. *Australian journal of physics*, pages 371–404 (1985).

[53] N. Stribeck, *X-Ray Scattering of Soft Matter*, book chapter Interpretation of Scattering Patterns, pages 95–190, Springer Berlin Heidelberg, Berlin, Heidelberg (2007), ISBN 978-3-540-69856-2, doi:10.1007/978-3-540-69856-2_8.

[54] R. de L. Kronig, On the Theory of Dispersion of X-Rays. *Journal of the Optical Society of America*, **12**(6):547–557 (1926), doi:10.1364/JOSA.12.000547.

[55] H. B. Stuhrmann, Anomalous small angle scattering. *Quarterly Reviews of Biophysics*, **14**(3):433–60 (1981).

[56] N. Miyamoto, N. Sakai, T. Hirayama, K. Miwa, Y. Oguro, H. Oki, K. Okada, T. Takagi, H. Iwata, Y. Awazu, S. Yamasaki, T. Takeuchi, H. Miki, A. Hori, and S. Imamura, Discovery of N-[5-(2-[(cyclopropylcarbonyl)amino]imidazo[1,2-b]pyridazin-6-yloxy)-2-methylphenyl]-1,3-dimethyl-1H-pyrazole-5-carboxamide (TAK-593), a highly potent VEGFR2 kinase inhibitor. *Bioorganic & Medicinal Chemistry*, **21**(8):2333–2345 (2013), doi:https://doi.org/10.1016/j.bmc.2013.01.074.

[57] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, The Protein Data Bank. *Nucleic Acids Research*, **28**(1):235–242 (2000), doi:10.1093/nar/28.1.235.

[58] University of Southern California - James Stroud, MakeNA server (Online; accessed 08-01-2018), URL `http://structure.usc.edu/make-na/server.html`.

[59] D. Svergun, C. Barberato, and M. H. Koch, CRYSOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography*, **28**:768–773 (1995), doi:10.1107/S0021889895007047.

[60] P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. J. Koch, and D. I. Svergun, PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*, **36**(5):1277–1282 (2003), doi:10.1107/S0021889803012779.

[61] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, **25**(13):1605–1612 (2004), doi:10.1002/jcc.20084.

[62] P. Debye and A. M. Bueche, Scattering by an Inhomogeneous Solid. *Journal of Applied Physics*, **20**(6):518–525 (1949), doi:10.1063/1.1698419.

[63] G. Porod, Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen. *Kolloid-Zeitschrift and Zeitschrift Fur Polymere*, **124**(2):83–114 (1951).

[64] O. Glatter, A new method for the evaluation of small-angle scattering data. *Journal of Applied Crystallography*, **10**(5):415–421 (1977), doi:10.1107/S0021889877013879.

[65] O. Glatter and O. Kratky, *Small Angle X-ray Scattering*, Academic Press London (1982), ISBN 0122862805.

[66] D. I. Svergun and M. H. J. Koch, Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics*, **66**:1735–1782 (2003), doi:10.1088/0034-4885/66/10/R05.

[67] D. A. Jacques and J. Trewhella, Small-angle scattering for structural biology—Expanding the frontier while avoiding the pitfalls. *Protein Science*, **19**(4):642–657 (2010), doi:10.1002/pro.351.

[68] M. H. Koch, P. Vachette, and D. I. Svergun, Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Quarterly Reviews of Biophysics*, **36**(2):147–227 (2003).

[69] R. S. Mathew-Fenn, R. Das, J. A. Silverman, P. A. Walker, and P. A. Harbury, A molecular ruler for measuring quantitative distance distributions. *PLoS One*, **3**(10):e3229 (2008), doi:10.1371/journal.pone.0003229.

[70] J. Lipfert and S. Doniach, Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu. Rev. Biophys. Biomol. Struct.*, **36**:307–327 (2007), doi:10.1146/annurev.biophys.36.040306.132655.

[71] K. Stovgaard, C. Andreetta, J. Ferkinghoff-Borg, and T. Hamelryck, Calculation of accurate small angle X-ray scattering curves from coarse-grained protein models. *BMC Bioinformatics*, **11**:429 (2010), doi:10.1186/1471-2105-11-429.

[72] M. B. Kozin and D. I. Svergun, Automated matching of high- and low-resolution structural models. *Journal of Applied Crystallography*, **34**:33–41 (2001).

[73] M. A. Wilson and A. T. Brunger, The 1.0 Å crystal structure of Ca2+-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. *Journal of Molecular Biology*, **301**(5):1237–1256 (2000), doi:10.1006/jmbi.2000.4029.

[74] V. V. Volkov and D. I. Svergun, Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography*, **36**(3):860–864 (2003), doi:10.1107/S0021889803000268.

[75] W. Wriggers, R. A. Milligan, and J. A. McCammon, Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *Journal of Structural Biology*, **125**(2-3):185–195 (1999).

[76] W. Wriggers and P. Chacon, Using Situs for the registration of protein structures with low-resolution bead models from X-ray solution scattering. *Journal of Applied Crystallography*, **34**:773–776 (2001).

[77] W. Hoppe, A New X-Ray Method for the Determination of the Quaternary Structure of Protein Complexes. *Israel Journal of Chemistry*, **10**(2):321–333 (1972), doi:10.1002/ijch.197200036.

[78] W. Hoppe, The label triangulation method and the mixed isomorphous replacement principle. *Journal of Molecular Biology*, **78**(3):581–585 (1973), doi:https://doi.org/10.1016/0022-2836(73)90480-4.

[79] O. Kratky and W. Worthmann, Über die Bestimmbarkeit der Konfiguration gelöster organischer Moleküle durch interferometrische Vermessung mit Röntgenstrahlen. *Monatshefte für Chemie und verwandte Teile anderer Wissenschaften*, **76**(3):263–281 (1947), doi:10.1007/bf00898977.

[80] O. Kratky, G. Porod, and A. Sekora, Zur röntgenographischen Untersuchung der Verknäuelung gelöster Fadenmoleküle. *Monatshefte für Chemie und verwandte Teile anderer Wissenschaften*, **78**(3):295–296 (1948), doi:10.1007/bf01141535.

[81] B. K. Vainshtein, L. A. Feigin, Y. M. Lvov, R. I. Gvozdev, S. A. Marakushev, and G. I. Likhtenshtein, Determination of the distance between heavy-atom markers in haemoglobin and histidine decarboxylase in solution by small-angle X-ray scattering. *FEBS Letters*, **116**(1):107–110 (1980), doi:10.1016/0014-5793(80)80539-4.

[82] R. S. Mathew-Fenn, R. Das, and P. A. Harbury, Remeasuring the double helix. *Science*, **322**:446–449 (2008), doi:10.1126/science.1158881.

[83] X. Shi, S. Bonilla, D. Herschlag, and P. Harbury, Quantifying Nucleic Acid Ensembles with X-ray Scattering Interferometry. *Methods in Enzymology*, **558**:75–97 (2015), doi:10.1016/bs.mie.2015.02.001.

[84] P. Willmott, *An Introduction to Synchrotron Radiation: Techniques and Applications*, book chapter Introduction, pages 1–14, John Wiley & Sons, Ltd, Chichester, UK (2011), ISBN 9781119970958, doi:10.1002/9781119970958.ch1.

[85] J. Als-Nielsen and D. McMorrow, *Elements of Modern X-ray Physics*, book chapter Sources, pages 29–67, John Wiley & Sons, Inc., Hoboken, NJ, USA (2011), ISBN 9781119998365, doi:10.1002/9781119998365.ch2.

[86] S. Doniach, Fourth-generation X-ray sources: some possible applications to biology. *Journal of Synchrotron Radiation*, **7**(Pt 3):116–20 (2000), doi:S0909049500004143[pii]10.1107/S0909049500004143.

[87] J. Cesar da Silva, A. Pacureanu, Y. Yang, S. Bohic, C. Morawe, R. Barrett, and P. Cloetens, Efficient concentration of high-energy x-rays for diffraction-limited imaging resolution. *Optica*, **4**(5):492–495 (2017), doi:10.1364/OPTICA.4.000492.

[88] M. Sztucki, E. Di Cola, and T. Narayanan, Instrumental developments for anomalous small-angle X-ray scattering from soft matter systems. *Journal of Applied Crystallography*, **43**(6):1479–1487 (2010).

[89] P. Ball, Europe's X-ray laser fires up. *Nature*, **548**:507–508 (2017).

[90] European X-ray free-electron laser (European XFEL), European XFEL Brilliance (Online; accessed 21-12-2017), URL `https://www.xfel.eu/facility/comparison/index_eng.html`.

[91] P. Willmott, *An Introduction to Synchrotron Radiation: Techniques and Applications*, book chapter Synchrotron Physics, pages 39–86, John Wiley & Sons, Ltd, Chichester, UK (2011), ISBN 9781119970958, doi:10.1002/9781119970958.ch3.

[92] A. Balerna and S. Mobilio, *Synchrotron Radiation: Basics, Methods and Applications*, book chapter Introduction to Synchrotron Radiation, pages 3–28, Springer Berlin Heidelberg, Berlin, Heidelberg (2015), ISBN 978-3-642-55315-8, doi:10.1007/978-3-642-55315-8_1.

[93] European Synchrotron Radiation Facility (ESRF), ESRF storage ring (Online; accessed: 03-01-2018), URL `http://www.esrf.eu/home/UsersAndScience/Accelerators/parameters.html`.

[94] P. Willmott, *An Introduction to Synchrotron Radiation: Techniques and Applications*, book chapter Beamlines, pages 87–131, John Wiley & Sons, Ltd, Chichester, UK (2011), ISBN 9781119970958, doi:10.1002/9781119970958.ch4.

[95] G. Margaritondo, *Synchrotron Radiation: Basics, Methods and Applications*, book chapter Characteristics and Properties of Synchrotron Radiation, pages 29–63, Springer Berlin Heidelberg, Berlin, Heidelberg (2015), ISBN 978-3-642-55315-8, doi:10.1007/978-3-642-55315-8_2.

[96] European Synchrotron Radiation Facility (ESRF) - Beamline BM29, X-ray source BM29 (Online; accessed: 03-01-2018), URL `http://www.esrf.eu/home/UsersAndScience/Experiments/MX/About_our_beamlines/bm29/beamline-specifications.html`.

[97] Stanford Synchrotron Radiation Source (SSRL) - Beamline 4-2, X-ray source BL 4-2 (Online; accessed: 20-01-2018), URL `https://www-ssrl.slac.stanford.edu/~saxs/instrumentation/instrumentation.htm`.

[98] Advanced Photon Source (APS) - Beamline 12-ID, X-ray source 12-ID (Online; accessed: 20-01-2018), URL `https://www.aps.anl.gov/Beamlines/Directory/Details?beamline_id=110`.

[99] G. Aquilanti, L. Vaccari, J. R. Plaisier, and A. Goldoni, *Synchrotron Radiation: Basics, Methods and Applications*, book chapter Instrumentation at Synchrotron Radiation Beamlines, pages 65–104, Springer Berlin Heidelberg, Berlin, Heidelberg (2015), ISBN 978-3-642-55315-8, doi:10.1007/978-3-642-55315-8_3.

[100] W. Bragg and W. Bragg, The reflection of X-rays by crystals. *Proceedings of the Royal Society of London. Series A*, **88**(605):428–438 (1913), doi:10.1098/rspa.1913.0040.

[101] R. P. Feynman, R. B. Leighton, and M. Sands, The Feynman Lectures on Physics. *The Feynman Lectures on Physics* (1989).

[102] N. Fischer, A. L. Konevega, W. Wintermeyer, M. V. Rodnina, and H. Stark, Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature*, **466**(7304):329–33 (2010), doi:10.1038/nature09206.

[103] H. Frauenfelder, New looks at protein motions. *Nature*, **338** (1989).

[104] J. P. Müller, A. Löf, S. Mielke, T. Obser, L. K. Brützel, W. Vanderlinden, J. Lipfert, R. Schneppenheim, and M. Benoit, pH-Dependent Interactions in Dimers Govern the Mechanics and Structure of von Willebrand Factor. *Biophysical Journal*, **111**(2):312–22 (2016), doi:10.1016/j.bpj.2016.06.022.

[105] X. Shi, D. Herschlag, and P. A. Harbury, Structural ensemble and microscopic elasticity of freely diffusing DNA by direct measurement of fluctuations. *Proceedings of the National Academy of Sciences of the United States of America*, **110**(16):E1444–E1451 (2013), doi:10.1073/pnas.1218830110.

[106] X. Shi, K. A. Beauchamp, P. B. Harbury, and D. Herschlag, From a structural average to the conformational ensemble of a DNA bulge. *Proceedings of the National Academy of Sciences of the United States of America*, **111**(15):E1473–E1480 (2014), doi:10.1073/pnas.1317032111.

[107] X. Shi, L. Huang, D. M. Lilley, P. B. Harbury, and D. Herschlag, The solution structural ensembles of RNA kink-turn motifs and their protein complexes. *Nature Chemical Biology*, **12**(January):146–152 (2016), doi:10.1038/nchembio.1997.

[108] X. Shi, P. Walker, P. A. Harbury, and D. Herschlag, Determination of the conformational ensemble of the TAR RNA by X-ray scattering interferometry. *Nucleic Acids Research*, **45**(8):e64 (2017), doi:10.1093/nar/gkw1352.

[109] C. J. Ackerson, M. T. Sykes, and R. D. Kornberg, Defined DNA nanoparticle conjugates. *Proceedings of the National Academy of Sciences of the United States of America*, **102**:13383–13385 (2005).

[110] A. P. Alivisatos, K. P. Johnsson, X. Peng, T. E. Wilson, C. J. Loweth, M. P. Bruchez, and P. G. Schultz, Organization of 'nanocrystal molecules' using DNA. *Nature*, **382**:609–11 (1996), doi:10.1038/382609a0.

[111] M. E. Aubin-Tam and K. Hamad-Schifferli, Gold nanoparticle-cytochrome c complexes: The effect of nanoparticle ligand charge on protein structure. *Langmuir*, **21**:12080–12084 (2005), doi:10.1021/la052102e.

[112] M. E. Aubin-Tam, W. Hwang, and K. Hamad-Schifferli, Site-directed nanoparticle labeling of cytochrome c. *Proceedings of the National Academy of Sciences of the United States of America*, **106**:4095–4100 (2009), doi:10.1073/pnas.0807299106.

[113] M. Azubel and R. D. Kornberg, Synthesis of Water-Soluble, Thiolate-Protected Gold Nanoparticles Uniform in Size. *Nano Letters*, **16**(5):3348–51 (2016), doi:10.1021/acs.nanolett.6b00981.

[114] T. G. Schaaff, G. Knight, M. N. Shafigullin, R. F. Borkman, and R. L. Whetten, Isolation and Selected Properties of a 10.4 kDa Gold:Glutathione Cluster Compound. *Journal of Physical Chemistry B*, **102**(52):10643–10646 (1998), doi:10.1021/jp9830528.

[115] Integrated Device Technology (IDT), IDT OligoAnalyzer 3.1 (Online; accessed 02-10-2017), URL `https://eu.idtdna.com/calc/analyzer`.

[116] A. Grishaev, *Sample Preparation, Data Collection, and Preliminary Data Analysis in Biomolecular Solution X-Ray Scattering*, volume Current Protocols in Protein Science, book section 17.14, pages 1–18, John Wiley & Sons, Inc. (2001), ISBN 9780471140863, doi:10.1002/0471140864.ps1714s70.

[117] J. E. Burke and S. E. Butcher, *Nucleic Acid Structure Characterization by Small Angle X-Ray Scattering (SAXS)*, volume Current Protocols in Nucleic Acid Chemistry, book section 7.18, pages 1–18, John Wiley & Sons, Inc. (2001), ISBN 9780471142706, doi:10.1002/0471142700.nc0718s51.

[118] S. Doniach and J. Lipfert, Small and wide angle x-ray scattering from biological macromolecules and their complexes in solution. *Comprehensive Biophysics*, **1**:376–397 (2012), doi:10.1016/B978-0-12-374920-8.00122-3.

[119] K. N. Dyer, M. Hammel, R. P. Rambo, S. E. Tsutakawa, I. Rodic, S. Classen, J. A. Tainer, and G. L. Hura, High-Throughput SAXS for the Characterization of Biomolecules in Solution: A Practical Approach. *Methods in molecular biology (Clifton, N.J.)*, **1091**:245–258 (2014), doi:10.1007/978-1-62703-691-7_18.

[120] G. L. Hura, H. Budworth, K. N. Dyer, R. P. Rambo, M. Hammel, C. T. McMurray, and J. A. Tainer, Comprehensive macromolecular conformations mapped by quantitative SAXS analyses. *Nature Methods*, **10**:453 (2013), doi:10.1038/nmeth.2453.

[121] C. M. Jeffries, M. A. Graewert, C. E. Blanchet, D. B. Langley, A. E. Whitten, and D. I. Svergun, Preparing Monodisperse Macromolecular Samples for Successful Biological Small-Angle X-ray and Neutron Scattering Experiments. *Nature protocols*, **11**(11):2122–2153 (2016), doi:10.1038/nprot.2016.113.

[122] A. T. Tuukkanen, A. Spilotros, and D. I. Svergun, Progress in small-angle scattering from biological solutions at high-brilliance synchrotrons. *IUCrJ*, **4**(Pt 5):518–528 (2017), doi:10.1107/S2052252517008740.

[123] Biological Small Angle Scattering Group, European Molecular Biology Laboratory (EMBL), Small Angle Scattering Biological Data Bank (SASBDB) (Online; accessed 02-10-2017), URL `https://www.sasbdb.org/`.

[124] B. R. Pauw, Everything SAXS: small-angle scattering pattern collection and correction. *Journal of Physics: Condensed Matter*, **26**:239501 (2014), doi:10.1088/0953-8984/26/23/239501.

[125] R. C. Miake-Lye, S. Doniach, and K. O. Hodgson, Anomalous x-ray scattering from terbium-labeled parvalbumin in solution. *Biophysical Journal*, **41**(3):287–92 (1983), doi:S0006-3495(83)84440-3[pii]10.1016/S0006-3495(83)84440-3.

[126] D. Arnlund, L. C. Johansson, C. Wickstrand, A. Barty, G. J. Williams, E. Malmerberg, J. Davidsson, D. Milathianaki, D. P. DePonte, R. L. Shoeman, D. Wang, D. James, G. Katona, S. Westenhoff, T. A. White, A. Aquila, S. Bari, P. Berntsen, M. Bogan, T. B. van Driel, R. B. Doak, K. S. Kjaer, M. Frank, R. Fromme, I. Grotjohann, R. Henning, M. S. Hunter, R. A. Kirian, I. Kosheleva, C. Kupitz, M. Liang, A. V. Martin, M. M. Nielsen, M. Messerschmidt, M. M. Seibert, J. Sjohamn, F. Stellato, U. Weierstall, N. A. Zatsepin, J. C. Spence, P. Fromme, I. Schlichting, S. Boutet, G. Groenhof, H. N. Chapman, and R. Neutze, Visualizing a protein quake with time-resolved X-ray scattering at a free-electron laser. *Nature Methods*, **11**(9):923–6 (2014), doi:10.1038/nmeth.3067.

[127] M. Bada, D. Walther, B. Arcangioli, S. Doniach, and M. Delarue, Solution structural studies and low-resolution model of the Schizosaccharomyces pombe sap1 protein. *Journal of Molecular Biology*, **300**(3):563–74 (2000), doi:10.1006/jmbi.2000.3854S0022-2836(00)93854-3[pii].

[128] D. Mendez, T. J. Lane, J. Sung, J. Sellberg, C. Levard, H. Watkins, A. E. Cohen, M. Soltis, S. Sutton, J. Spudich, V. Pande, D. Ratner, and S. Doniach, Observation of correlated X-ray scattering at atomic resolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **369**(1647):20130315 (2014), doi:10.1098/rstb.2013.0315.

[129] D. Mendez, H. Watkins, S. Qiao, K. S. Raines, T. J. Lane, G. Schenk, G. Nelson, G. Subramanian, K. Tono, Y. Joti, M. Yabashi, D. Ratner, and S. Doniach, Angular correlations of photons from solution diffraction at a free-electron laser encode molecular structure. *IUCrJ*, **3**(Pt 6):420–429 (2016), doi:10.1107/S2052252516013956.

[130] G. Schenk, B. Krajina, A. Spakowitz, and S. Doniach, Potential for measurement of the distribution of DNA folds in complex environments using Correlated X-ray Scattering. *Modern physics letters. B, Condensed matter physics, statistical physics, applied physics*, **30**(8):1650117 (2016), doi:10.1142/S0217984916501177.

[131] SLAC and LCLS, Building X-ray lasers. *Nature Photonics*, **4**:0–1 (2010), doi:10.1038/nphoton.2010.276.

[132] I. L. Smolsky, P. Liu, M. Niebuhr, K. Ito, T. M. Weiss, and H. Tsuruta, Biological small-angle x-ray scattering facility at the Stanford synchrotron radiation laboratory. *Journal of Applied Crystallography*, **40**:S453–S458 (2007), doi:10.1107/s0021889807009624.

[133] A. Andrus and R. G. Kuimelis, *Polyacrylamide Gel Electrophoresis (PAGE) of Synthetic Nucleic Acids*, volume Current Protocols in Nucleic Acid Chemistry, book section 10.4, pages 1–10, John Wiley & Sons, Inc. (2001), ISBN 9780471142706, doi:10.1002/0471142700.nc1004s01.

[134] N. D. Sinha and K. E. Jung, *Analysis and Purification of Synthetic Nucleic Acids Using HPLC*, volume Current Protocols in Nucleic Acid Chemistry, book section 10.5, pages 1–39, John Wiley & Sons, Inc. (2001), ISBN 9780471142706, doi:10.1002/0471142700.nc1005s61.

[135] S. M. Müller, H. Galliardt, J. Schneider, B. G. Barisas, and T. Seidel, Quantification of Förster resonance energy transfer by monitoring sensitized emission in living plant cells. *Frontiers in Plant Science*, **4**:1–20 (2013), doi:10.3389/fpls.2013.00413.

[136] Z. Yang, G. Jimènez-Osès, C. J. Lòpez, M. D. Bridges, K. N. Houk, and W. L. Hubbell, Long-range distance measurements in proteins at physiological temperatures using saturation recovery EPR spectroscopy. *Journal of the American Chemical Society*, **136**:15356–15365 (2014), doi:10.1021/ja5083206.

[137] G. Krainer, A. Hartmann, and M. Schlierf, farFRET: Extending the Range in Single-Molecule FRET Experiments beyond 10 nm. *Nano Letters*, **15**(9):5826–9 (2015), doi:10.1021/acs.nanolett.5b01878.

[138] A. J. Mastroianni, D. A. Sivak, P. L. Geissler, and A. P. Alivisatos, Probing the conformational distributions of subpersistence length DNA. *Biophysical Journal*, **97**(5):1408–1417 (2009), doi:10.1016/j.bpj.2009.06.031.

[139] V. J. Pinfield and D. J. Scott, Anomalous small angle X-ray scattering simulations: proof of concept for distance measurements for nanoparticle-labelled biomacromolecules in solution. *PLoS One*, **9**(4):e95664 (2014), doi:10.1371/journal.pone.0095664.

[140] M. A. Beno, G. Jennings, M. Engbretson, G. S. Knapp, C. Kurtz, B. Zabransky, J. Linton, S. Seifert, C. Wiley, and P. A. Montano, Basic Energy Sciences Synchrotron Radiation Center Undulator Sector at the Advanced Photon Source. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **467–468, Part 1**:690–693 (2001), doi:http://dx.doi.org/10.1016/S0168-9002(01)00455-7.

[141] R. G. Munro, Reliability of partial structure factors determined by anomalous dispersion of x-rays. *Physical Review B: Condensed Matter Physics*, **25**(8):5037–5045 (1982), doi:10.1103/PhysRevB.25.5037.

[142] M. Saito and Y. Waseda, Anomalous X-ray scattering for determining the partial structural functions of binary liquids. *Journal of Synchrotron Radiation*, **7**(3):152–159 (2000).

[143] J. M. Schurr, A Possible Cooperative Structural Transition of DNA in the 0.25–2.0 pN Range. *Journal of Physical Chemistry B*, **119**:6389–6400 (2015), doi:10.1021/acs.jpcb.5b03174.

[144] R. Das, T. T. Mills, L. W. Kwok, G. S. Maskel, I. S. Millett, S. Doniach, K. D. Finkelstein, D. Herschlag, and L. Pollack, Counterion distribution around DNA probed by solution X-ray scattering. *Physical Review Letters*, **90**(18):188103 (2003).

[145] S. A. Pabit, K. D. Finkelstein, and L. Pollack, Using anomalous small angle X-ray scattering to probe the ion atmosphere around nucleic acids. *Methods in Enzymology*, **469**:391–410 (2009), doi:S0076-6879(09)69019-4[pii]10.1016/S0076-6879(09)69019-4.

[146] R. Holliday, A mechanism for gene conversion in fungi. *Genetical Research*, **5**(2):282–304 (1964), doi:10.1017/S0016672300001233.

[147] H. Potter and D. Dressler, On the mechanism of genetic recombination: electron microscopic observation of recombination intermediates. *Proceedings of the National Academy of Sciences of the United States of America*, **73**:3000–3004 (1976).

[148] N. C. Seeman and N. R. Kallenbach, Design of immobile nucleic acid junctions. *Biophysical Journal*, **44**:201–9 (1983), doi:10.1016/S0006-3495(83)84292-1.

[149] M. M. Cox, M. F. Goodman, K. N. Kreuzer, D. J. Sherratt, S. J. Sandler, and K. J. Marians, The importance of repairing stalled replication forks. *Nature*, **404**:37 (2000), doi:10.1038/35003501.

[150] D. M. Lilley, Structures of helical junctions in nucleic acids. *Quarterly Reviews of Biophysics*, **33**(2):109–59 (2000).

[151] N. C. Seeman, DNA in a material world. *Nature*, **421**(6921):427–31 (2003), doi:10.1038/nature01406.

[152] R. J. Grainger, A. I. H. Murchie, and D. M. J. Lilley, Exchange between Stacking Conformers in a Four-Way DNA Junction. *Biochemistry*, **37**(1):23–32 (1998), doi:10.1021/bi9721492.

[153] M. F. White and D. M. J. Lilley, The Structure-selectivity and Sequence-preference of the Junction-resolving Enzyme CCE1 ofSaccharomyces cerevisiae. *Journal of Molecular Biology*, **257**(2):330–341 (1996), doi:https://doi.org/10.1006/jmbi.1996.0166.

[154] D. M. Lilley and M. F. White, The junction-resolving enzymes. *Nature Reviews. Molecular Cell Biology*, **2**(6):433–443 (2001), doi:10.1038/35073057.

[155] S. A. McKinney, A.-C. Déclais, D. M. J. Lilley, and T. Ha, Structural dynamics of individual Holliday junctions. *Nature Structural Biology*, **10**:93–97 (2003), doi:10.1038/nsb883.

[156] D. M. J. Lilley, Holliday junction-resolving enzymes—structures and mechanisms. *FEBS Letters*, **591**(8):1073–1082 (2017), doi:10.1002/1873-3468.12529.

[157] D. R. Duckett, A. I. H. Murchie, S. Diekmann, E. von Kitzing, B. Kemper, and D. M. J. Lilley, The structure of the holliday junction, and its resolution. *Cell*, **55**:79–89 (1988), doi:10.1016/0092-8674(88)90011-6.

[158] J. P. Cooper and P. J. Hagerman, Geometry of a branched DNA structure in solution. *Proceedings of the National Academy of Sciences of the United States of America*, **86**:7336–40 (1989), doi:10.1073/pnas.86.19.7336.

[159] C. Joo, S. A. McKinney, D. M. Lilley, and T. Ha, Exploring rare conformational species and ionic effects in DNA Holliday junctions using single-molecule spectroscopy. *Journal of Molecular Biology*, **341**(3):739–51 (2004), doi:10.1016/j.jmb.2004.06.024.

[160] S. Hohng, R. Zhou, M. K. Nahas, J. Yu, K. Schulten, D. M. J. Lilley, and T. Ha, Fluorescence-Force Spectroscopy Maps Two-Dimensional Reaction Landscape of the Holliday Junction. *Science*, **318**:279–283 (2007).

[161] D. R. Duckett, A. I. Murchie, and D. M. Lilley, The role of metal ions in the conformation of the four-way DNA junction. *The EMBO Journal*, **9**(2):583–590 (1990).

[162] J. E. Mueller, B. Kemper, R. P. Cunningham, N. R. Kallenbach, and N. C. Seeman, T4 endonuclease VII cleaves the crossover strands of Holliday junction analogs. *Proceedings of the National Academy of Sciences of the United States of America*, **85**(24):9441–9445 (1988), doi:10.1073/pnas.85.24.9441.

[163] R. J. Bennett and S. C. West, RuvC protein resolves Holliday junctions via cleavage of the continuous (noncrossover) strands. *Proceedings of the National Academy of Sciences of the United States of America*, **92**(12):5635–5639 (1995).

[164] D. N. Gopaul, F. Guo, and G. D. Van Duyne, Structure of the Holliday junction intermediate in Cre-*loxP* site-specific recombination. *The EMBO Journal*, **17**:4175–4187 (1998).

[165] D. M. Lilley and D. G. Norman, The Holliday junction is finally seen with crystal clarity. *Nature Structural Biology*, **6**:897–9 (1999), doi:10.1038/13255.

[166] A. L. Mikheikin, A. Y. Lushnikov, and Y. L. Lyubchenko, Effect of DNA supercoiling on the geometry of holliday junctions. *Biochemistry*, **45**:12998–13006 (2006), doi:10.1021/bi061002k.

[167] M. Nöllmann, W. M. Stark, and O. Byron, Low-resolution reconstruction of a synthetic DNA holliday junction. *Biophysical Journal*, **86**:3060–9 (2004), doi:10.1016/S0006-3495(04)74355-6.

[168] J. Wang, P. Daldrop, L. Huang, and D. M. J. Lilley, The k-junction motif in RNA structure. *Nucleic Acids Research*, **42**(8):5322–5331 (2014), doi:10.1093/nar/gku144.

[169] E. V. Kitzing, D. M. J. Lilley, and S. Diekmann, The stereochemistry of theoretical study a four-way DNA junction : a theoretical study. *Nucleic Acids Research*, **18**:2671–2683 (1990).

[170] R. M. Clegg, A. I. Murchie, A. Zechel, C. Carlberg, S. Diekmann, and D. M. Lilley, Fluorescence resonance energy transfer analysis of the structure of the four-way DNA junction. *Biochemistry*, **31**:4846–4856 (1992), doi:10.1021/bi00135a016.

[171] M. Greenfeld and D. Herschlag, Fluorescently labeling synthetic RNAs. *Methods in Enzymology*, **530**:281–97 (2013), doi:10.1016/B978-0-12-420037-1.00015-4.

[172] S. Bonilla, C. Limouse, N. Bisaria, M. Gebala, H. Mabuchi, and D. Herschlag, Single-Molecule Fluorescence Reveals Commonalities and Distinctions among Natural and in Vitro-Selected RNA Tertiary Motifs in a Multistep Folding Pathway. *Journal of the American Chemical Society*, **139**(51):18576–18589 (2017), doi:10.1021/jacs.7b08870.

[173] M. Greenfeld, D. S. Pavlichin, H. Mabuchi, and D. Herschlag, Single molecule analysis research tool (SMART): An integrated approach for analyzing single molecule data. *PLoS ONE*, **7** (2012), doi:10.1371/journal.pone.0030024.

[174] J. Lipfert, R. Das, V. B. Chu, M. Kudaravalli, N. Boyd, D. Herschlag, and S. Doniach, Structural transitions and thermodynamics of a glycine-dependent riboswitch from Vibrio cholerae. *Journal of Molecular Biology*, **365**(5):1393–406 (2007), doi:S0022-2836(06)01360-X[pii]10.1016/j.jmb.2006.10.022.

[175] J. Lipfert, J. Ouellet, D. G. Norman, S. Doniach, and D. M. Lilley, The complete VS ribozyme in solution studied by small-angle X-ray scattering. *Structure*, **16**(9):1357–1367 (2008), doi:S0969-2126(08)00293-1[pii]10.1016/j.str.2008.07.007.

[176] R. P. Rambo and J. A. Tainer, Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. *Current Opinion in Structural Biology*, **20**(1):128–137 (2010), doi:10.1016/j.sbi.2009.12.015.

[177] R. P. Rambo and J. A. Tainer, Improving small-angle X-ray scattering data for structural analyses of the RNA world. *RNA*, **16**(3):638–646 (2010), doi:10.1261/rna.1946310.

[178] R. P. Rambo and J. A. Tainer, Super-Resolution in Solution X-Ray Scattering and Its Applications to Structural Systems Biology. *Annu Rev Biophys*, **42**:415–441 (2013), doi:10.1146/annurev-biophys-083012-130301.

[179] W. A. Cantara, E. D. Olson, and K. Musier-Forsyth, Analysis of RNA structure using small-angle X-ray scattering. *Methods*, **113**:46–55 (2017), doi:http://dx.doi.org/10.1016/j.ymeth.2016.10.008.

[180] Y. R. Bhandari, W. Jiang, E. A. Stahlberg, J. R. Stagno, and Y.-X. Wang, Modeling RNA topological structures using small angle X-ray scattering. *Methods*, **103**:18–24 (2016), doi:http://dx.doi.org/10.1016/j.ymeth.2016.04.015.

[181] M. Ali, J. Lipfert, S. Seifert, D. Herschlag, and S. Doniach, The ligand-free state of the TPP riboswitch: a partially folded RNA structure. *Journal of Molecular Biology*, **396**(1):153–65 (2010), doi:S0022-2836(09)01407-7[pii]10.1016/j.jmb.2009.11.030.

[182] D. Klein, T. Schmeing, P. Moore, and T. Steitz, The kink-turn: a new RNA secondary structure motif. *The EMBO Journal*, **20**(15):4214–4221 (2001), doi:10.1093/emboj/20.15.4214.

[183] J. Stigler and M. Rief, Calcium-dependent folding of single calmodulin molecules. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(44):17814–17819 (2012), doi:10.1073/pnas.1201801109.

[184] B. A. Seaton, J. F. Head, D. M. Engelman, and F. M. Richards, Calcium-induced increase in the radius of gyration and maximum dimension of calmodulin measured by small-angle x-ray scattering. *Biochemistry*, **24**(24):6740–6743 (1985), doi:10.1021/bi00345a002.

[185] E. W. Small and S. R. Anderson, Fluorescence anisotropy decay demonstrates calcium-dependent shape changes in photo-cross-linked calmodulin. *Biochemistry*, **27**(1):419–428 (1988), doi:10.1021/bi00401a063.

[186] G. Barbato, M. Ikura, L. E. Kay, R. W. Pastor, and A. Bax, Backbone dynamics of calmodulin studied by nitrogen-15 relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry*, **31**(23):5269–5278 (1992), doi:10.1021/bi00138a005.

[187] J. J. Chou, S. Li, C. B. Klee, and A. Bax, Solution structure of Ca2+–calmodulin reveals flexible hand-like properties of its domains. *Nature Structural Biology*, **8**:990 (2001), doi:10.1038/nsb1101-990.

[188] N. J. Anthis, M. Doucleff, and G. M. Clore, Transient, Sparsely Populated Compact States of Apo and Calcium-Loaded Calmodulin Probed by Paramagnetic Relaxation Enhancement: Interplay of Conformational Selection and Induced Fit. *Journal of the American Chemical Society*, **133**(46):18966–18974 (2011), doi:10.1021/ja2082813.

[189] I. Bertini, C. Del Bianco, I. Gelis, N. Katsaros, C. Luchinat, G. Parigi, M. Peana, A. Provenzani, and M. A. Zoroddu, Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(18):6841 (2004).

[190] D. B. Heidorn and J. Trewhella, Comparison of the crystal and solution structures of calmodulin and troponin C. *Biochemistry*, **27**(3):909–915 (1988), doi:10.1021/bi00403a011.

[191] M. Kataoka, J. F. Head, T. Vorherr, J. Krebs, and E. Carafoli, Small-angle x-ray scattering study of calmodulin bound to two peptides corresponding to parts of the calmodulin-binding domain of the plasma membrane calcium pump. *Biochemistry*, **30**(25):6247–6251 (1991), doi:10.1021/bi00239a024.

[192] V. Majava, M. V. Petoukhov, N. Hayashi, P. Pirila, D. I. Svergun, and P. Kursula, Interaction between the C-terminal region of human myelin basic protein and calmodulin: analysis of complex formation and solution structure. *BMC Struct Biol*, **8**:10 (2008), doi:1472-6807-8-10[pii]10.1186/1472-6807-8-10.

[193] D. Schneidman-Duhovny, M. Hammel, and A. Sali, FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res*, **38**(Web Server issue):W540–4 (2010), doi:10.1093/nar/gkq461.

[194] L. S. Churchman, Z. Ökten, R. S. Rock, J. F. Dawson, and J. A. Spudich, Single molecule high-resolution colocalization of Cy3 and Cy5 attached to macromolecules measures intramolecular distances through time. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(5):1419 (2005).

[195] A. R. Dunn and J. A. Spudich, Dynamics of the unbound head during myosin V processive translocation. *Nature Structural & Molecular Biology*, **14**:246 (2007), doi:10.1038/nsmb1206https://www.nature.com/articles/nsmb1206#supplementary-information.

[196] P. Bernadó, E. Mylonas, M. V. Petoukhov, M. Blackledge, and D. I. Svergun, Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering. *Journal of the American Chemical Society*, **129**(17):5656–5664 (2007), doi:10.1021/ja069124n.

[197] D. Schneidman-Duhovny, M. Hammel, J. A. Tainer, and A. Sali, FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Research*, **44**(W1):W424–W429 (2016), doi:10.1093/nar/gkw389.

[198] M. Pelikan, G. L. Hura, and M. Hammel, Structure and flexibility within proteins as identified through small angle X-ray scattering. *General physiology and biophysics*, **28**(2):174–189 (2009).

[199] L. Zhang, D. Lei, J. M. Smith, M. Zhang, H. Tong, X. Zhang, Z. Lu, J. Liu, A. P. Alivisatos, and G. Ren, Three-dimensional structural dynamics and fluctuations of DNA-nanogold conjugates by individual-particle electron tomography. *Nature Communications*, **7**:11083 (2016), doi:10.1038/ncomms11083.

[200] P. D. Jadzinsky, G. Calero, C. J. Ackerson, D. A. Bushnell, and R. D. Kornberg, Structure of a Thiol Monolayer-Protected Gold Nanoparticle at 1.1 Å Resolution. *Science*, **318**:430–433 (2007).

[201] R. Gopalakrishna and W. B. Anderson, Ca^{2+}-induced hydrophobic site on calmodulin: Application for purification of calmodulin by phenyl-Sepharose affinity chromatography. *Biochemical and Biophysical Research Communications*, **104**(2):830–836 (1982), doi:http://dx.doi.org/10.1016/0006-291X(82)90712-4.

[202] S. Seifert, R. Winans, D. Tiede, and P. Thiyagarajan, Design and performance of a ASAXS instrument at the Advanced Photon Source. *Journal of Applied Crystallography*, **33**(3 Part 1):782–784 (2000), doi:doi:10.1107/S0021889800001333.

[203] B. Fleury, R. Cortes-Huerto, O. Taché, F. Testard, N. Menguy, and O. Spalla, Gold Nanoparticle Internal Structure and Symmetry Probed by Unified Small-Angle X-ray Scattering and X-ray Diffraction Coupled with Molecular Dynamics Analysis. *Nano Letters*, **15**(9):6088–6094 (2015), doi:10.1021/acs.nanolett.5b02924.

List of Figures

List of Tables

Acknowledgments

I would like to thank a number of people without whom this thesis would not have been possible:
Jan Lipfert for doing an amazing job supervising me and giving highly valuable feedback. I would also like to thank him for the great teamwork for numerous publications, grant proposals and projects, the relaxed and brilliant atmosphere in the group and various leisure activities. Thank you for the perfect balance of independence, trust and right advice. Of course also thanks for helping me to explore the scientific community during various conferences and making the research visit to Stanford possible.
Daniel Herschlag for hosting me in Stanford, his invitation to the workshop in Telluride and for the great collaboration. The time in his group was fantastic and also very productive. He really helped me with lots of valuable input and discussions about all sorts of problems.
Hermann Gaub for the genial and nice atmosphere at the entire chair as well as useful advice and discussions.
Many of the results wouldn't have been possible without the collaboration with **Pehr A. B. Harbury** and **Rhiju Das** at Stanford University as well as **Rebecca Waterman** at Harvard University.
Prof. Martin Zacharias for the great collaboration and the joined project.
Xuesong Shi for introducing me to all the experimental steps needed for XSI, his help and useful advice.
Sylvia Kreuzer for the great organization, nice conversations, keeping things running and making all the paperwork as smooth as possible. I would also like to thank **Thomas Nicolaus** for his assistance with many biochemical questions and **Angelika Kardinal** for her support in the lab.
Korbinian Liebl for the simulations, useful discussions and fun activities aside from scientific projects.
Steffen Sedlak for the teamwork at ESRF, discussions about scattering theory and being a great office mate. Of course also thanks for proofreading.
Steve Bonilla for his support with single molecule fluorescence sample preparation, measurements and the nice time in and outside the lab.
Philipp Walker for fun conversations, discussions and motivating me in the Phd and the BWL Kurs.
Franziska Kriegel for her help, feedback, discussions of all kinds during numerous coffee breaks and the motivating me.
Magnus Bauer for his support, contributing to the fun atmosphere in the office, great discussions and of course fruit tasting.
Margaux Pinney for being a great desk neighbor and fun conversations.
Willem Vanderlinden for fruitful discussions and lots of useful advice.
Linda Brützel for helpful discussions about small-angle scattering.
My office fellows **Leonard Schendel**, **Carleen Kluger**, **Larissa Zott** and **Sebastian Konrad**

for a nice time and fun conversations at the office.
Wolfgang Ott for providing me protein samples and his advice concerning biochemical questions. I would also like to thank **Lukas Milles** for useful discussions.
Thomas Carell for providing HPLC systems. Further, I would like to thank **Matthias Kurz** for his help operating the HPLCs and counsel to obtain the best results possible.
CeNS, **NIM** and **SFB** for organizing great events and providing me funding for conferences and research visits. I am also grateful to the **DAAD** for providing funding.
None of the results that lead to this thesis would have been possible without the members of **AG Lipfert**, **LS Gaub** and **Herschlag Lab**. Thank you for the great atmosphere, all the help and the nice time.
Finally, I would like to thank all the other people who have been supporting me through all these years. Special thanks to my parents **Patricia** and **Rainer**, my siblings **Michaela** and **Florian**, **Katherina** and my friends. I am thankful for your help and motivation making my studies and this work possible.